A Highly-Accurate Three-Way Decision-Incorporated Online Sparse Streaming Features Selection Model

Ruiyang Xu, *Graduate Student Member, IEEE*, Di Wu, *Member, IEEE*, Renfang Wang, and Xin Luo, *Fellow, IEEE*

***Abstract*—An online streaming feature selection (OSFS) model is highly efficient in processing the high-dimensional streaming features. In practical big data-related applications, streaming features are mostly highly-incomplete due to various unpredictable reasons like the privacy protection, leading to the issue of online sparse streaming feature selection (OS2FS). The incomplete streaming features can lead to the uncertain relationship between the labels and sparse features during the feature selection process, yet existing OSFS and OS2FS models focus on the certain relationships, resulting in accuracy loss by improperly-selected features. To address this critical issue, this article presents a three (3)-way decision-incorporated OS2FS (3WDO) model with the following two-fold ideas: 1) utilizing the latent factor analysis (LFA) approach to pre-estimate the missing data of the concerned sparse streaming features and 2) integrating the three-way decision (3WD) into the streaming features selection process for appropriately modeling the uncertainty within the label-feature interactions. By doing so, the uncertain relationships between labels and sparse features are characterized by more information and looser tolerance, thereby minimizing the decision risk of feature selection. Experimental results on twelve real-world datasets demonstrate that the proposed 3WDO model significantly outperforms seven state-of-the-art OSFS and OS2FS models, which strongly supports its ability of addressing practical issues.**

***Index Terms*—Latent factor analysis (LFA), online feature selection, sparse streaming features, three-way decision, uncertainty.**

Received 10 May 2024; revised 7 December 2024; accepted 26 February 2025. Date of publication 31 March 2025; date of current version 19 May 2025. This work was supported by the National Natural Science Foundation of China under Grant 62176070 and Grant 62272078. This article was recommended by Associate Editor Z. Liu. *(Corresponding authors: Renfang Wang; Xin Luo.)*

Ruiyang Xu is with the School of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing, 400065, and also with the Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences, Chongqing 400714, China (e-mail: d220201035@stu.cqupt.edu.cn).

Di Wu and Xin Luo are with the College of Computer and Information Science, Southwest University, Chongqing 400715, China (e-mail: wudi.cigit@gmail.com; luoxin21@gmail.com).

Renfang Wang is with the College of Big Data and Software Engineering, Zhejiang Wanli University, Ningbo 315100, China (e-mail: renfang_wang@126.com).

This article has supplementary material provided by the authors and color versions of one or more figures available at https://doi.org/10.1109/TSMC.2025.3548648.

Digital Object Identifier 10.1109/TSMC.2025.3548648

I. Introduction

With the rapid development of information technology, massing high-dimensional data are mostly encountered with multilevel, multigranularity, and multimodality, which set great challenges to related techniques like artificial intelligence, data management, communication, and storage [1], [2]. Feature selection is an effective approach to handling such high-dimensional data [3], [4], [5]. Recently, various feature selection models have sprung up [6], [7], including the filter-based [8], [9], wrapper-based [10], and embedded-based [11], [12] ones. Moreover, in big data-related applications, the feature space constantly explodes to even infinite [13], [14]. Thus, online streaming feature selection (OSFS) is put forward. For instance, Wu et al. [15] first proposed an OSFS model based on the online relevance and redundancy analyses. It divides the streaming features into strongly, weakly relevant and irrelevant groups to select strong and weak relevant but nonredundant features. Yu et al. [16] proposed a SAOLA model by generalizing the pairwise relationship of streaming features calculated by fly mutation.

However, existing OSFS models are constantly designed for the complete streaming features scenario, i.e., the target steaming features are complete without missing data. Nevertheless, in most real applications, the streaming features commonly have massive missing data due to various unpredictable reasons. For example, limited by the cell sequencing technology, it is difficult to analyze the measured values of all cells to assign them with a certain weight [17]. On the other hand, it is impossible to collect the complete patients data due to equipment limitations or human misunderstandings [18]. As a result, the issue of online sparse streaming feature selection (OS^2FS) emerges, i.e., how to perform accurate streaming feature selection when the feature stream is filled with numerous missing data?

In real-time recommendation systems, features such as user behavior logs and product attributes arrive in a streaming manner. Due to the fact that users cannot interact with all the products, missing data issues often arise. These features are interrelated with one another, making it challenging to clearly decide which features to retain and which to discard for comprehensively and accurately capturing user preferences. In this context, selecting the most informative subset becomes crucial for generating timely and accurate recommendations.

To achieve this, it is necessary to effectively pre-estimate the missing values and reduce the uncertainty in the feature selection process. To handle the incomplete streaming features, Wu et al. [19] innovated a framework for online sparse streaming feature selection, whose principle is to accurately estimate the missing streaming features before conducting the selection. However, the errors between the estimated and real data of sparse streaming features may cause uncertainty within the interactions between labels and sparse features [20]. Direct feature selection with the conventionally certain strategies may lead to accuracy and efficiency degeneration. Therefore, it is vital to minimize the effects of such uncertain relationships during the selection process.

The three-way decision is applied to the uncertain decision-making problem. The core idea is to divide the whole dataset into three subsets or parts (i.e., positive, boundary, and negative domain) by means of trisecting-acting-outcome (TAO) and adopt different decision-making strategies for different subsets or parts. Compared with the two-way decision, the three-way decision (3WD) introduces a delayed decision branch, i.e., putting the uncertain objects into the boundary domain [21]. It is necessary to reduce the influence of uncertainty of estimated data on feature selection. Inspired by the philosophy of the three-way decision, this study improves the 3WD theory to adapt to the OS^{2}FS problem. It proposes the three-way relevance analysis, which divides the estimated sparse streaming features into tri-partitions, i.e., the selecting (i.e., strong relevant), the delaying (i.e., weak relevant), and the discarding (i.e., irrelevant). The uncertain features (e.g., weakly-relevant features) can be processed as the delaying to wait for more information which maybe available during the subsequent process. As such, the decision risk of selecting the uncertain features can be significantly reduced. To this end, this article proposes a three (3)-way decision-incorporated OS^{2}FS (3WDO) model with the following two-fold ideas.

1) Adopting the latent factor analysis (LFA) approach to pre-estimate the missing data in the sparse streaming features.
2) Improving and utilizing the 3WD to delay the uncertain features, waiting for more information to facilitate the safe feature selection.

By doing so, this article makes the following contributions.

1) Innovatively proposing the 3WDO model with high accuracy and efficiency in estimating the sparse feature of OS^{2}FS;
2) It addresses the previous OS^{2}FS model's uncertain problems by integrating 3WD-based uncertain decision process;
3) Presenting the detailed technical principle, algorithm design, and rigorous theoretical analyses regarding the decision cost of the 3WDO model, which demonstrates that the three-way relevance analysis will reduce the decision cost; and
4) Conducting the extensive experiments on twelve real-world datasets for validating the performance of the 3WDO model. The experimental results illustrate that: 1) the 3WDO model significantly outperforms seven state-of-the-art OSFS and OS^{2}FS models in feature selection precision; 2) the 3WDO model has the best performance as the missing data ratio increases during OS^{2}FS; and 3) the 3WDO model minimizes the uncertain decision risk for OS^{2}FS,

Section II introduces the related work. Section III gives the preliminaries. Section IV presents the 3WDO model. Finally, Section VI concludes this article.

II. Related Work

An OSFS model effectively handles feature streams on the fly, which has attracted wide attention. Perkins and Theiler [22] build the regularized framework Grafting for online feature selection, which calls for the regularization parameters adjustments before selecting the related features. Therefore, it is not suitable for a circumstance with unknown feature space. The Alpha-investing presented by Zhou et al. [23] can deal with infinite feature streams but ignoring the redundant analysis of selected features. Wu et al. [15] divided features into strongly relevant, weakly relevant, and irrelevant ones to present two OSFS models, i.e., OSFS and Fast-OSFS, where the latter considers the redundancy between a new feature and already-selected ones. Yu et al. [16] proposed the SAOLA model based on mutual information, which only considers the pair-wise feature interactions. By utilizing the interactions between the streaming groups, Zhou et al. [24] presented the OGSFS-FI model. They further present the SFS-FI [25] model that is able to select features interacting with each other, including two-way, three-way, and N-way interactions. For appropriately modeling the dynamic decision, Zhou et al. [26] adopted the principle of 3WD to build the OSSFS-DD model that calculates the partition threshold following the 3WD theory for reducing the decision risk.

On the other hand, a rough set is a profitable tool for OSFS. Zhou et al. [27] presented the OFS-A3M model based on the neighborhood rough set relation affected by the adapted neighbors to select high relevance, high dependence, and low redundancy features. They subsequently present the OFS-Density [28] model that adopts a new adaptive density neighborhood relation for studying the domain information and setup parameter. Lou et al. [29] utilized the rough hypercuboid to present the RHDOFS model. Shu et al. [30] proposed an ANOHFS model based on adaptive neighborhood to effectively select the closely related feature hierarchy in high-dimensional data. The above mentioned methods, in spite of their efficiency in addressing the OSFS issue, are all focusing on the complete streaming features only but ignoring the missing features in OS^{2}FS scenarios.

For addressing the issue of OS^{2}FS efficiently, this article presents the 3WDO model with two-fold ideas: 1) estimating the unknown data in the concerned online sparse streaming features accurately and 2) improving and adopting the 3WD principle for reducing the decision risk, thus improving the selection results. To date, an LFA model is beneficial in missing data estimation [31]. It maps the known entries of the target high-dimensional and incomplete matrix to its row and column nodes [32], constructs a learning objective to judge the difference between the actual data and the estimated ones, and then builds the complete but low-rank approximation to the target incomplete matrix with the minimum generalized error measured by the learning objective [33], [34], [35], [36], [37], [38].

Besides, existing OS^2FS models adopt certain relevance and redundancy analysis to remove irrelevant and redundant features. An uncertain feature is wrongly classified as irrelevance, which cannot be recovered after being deleted, and the accuracy of the algorithm will be reduced. From the semantic explanations of positive, boundary, and negative regions used in the rough set, the 3WD is presented by Yao [39] to deal with complex and uncertain decision-making. Theoretically, a 3WD is better than a two-way decision [40].

With the development of the 3WD, the TAO model is formulated [41], [42], which consists of three steps: 1) dividing the whole set into three parts; 2) adopting a strategy according to the regional characteristics; and 3) evaluating the effects of trisecting and acting. Then, the 3WD is applied to the subsequent problem like classification to achieves the output results [5], [43], [46], [47]. Based on the above discussion, improving and applying the 3WD principle to address the issue of OS^2FS becomes feasible. Combining with the idea of 3WD, the division strategy is improved to make it suitable for three-way relevance analysis.

III. Preliminaries

A. Online Streaming Feature Selection

Table I summarizes the adopted symbols. OSFS selects the strongly and weakly relevant, but nonredundancy features on the fly. Given the streaming feature set $F = \{F_1, F_2, \ldots, F\psi_t\}$ as $F_t = [f_{1,t}, f_{2,t}, \ldots, f_{N,t}]^T$ with N instance at the timestamp t$\in\{1, 2, \ldots, \psi_t\}$. Conditional independence is defined as follows.

Definition 1 (Conditional Independence [15]): Given two distinct features F_t, $F_l \in F$, $t{\neq}l$, t, l$\in\{1, 2, \ldots, \psi_t\}$, they are conditionally independent on the subset $S\subseteq F$ with the conditions

of $P(F_t|F_l, S) = P(F_t, S)$ or $P(F_t|F_l, S) = P(F_l, S)$.

Thus, the strongly relevant, weakly relevant, and irrelevant feature of the class attribute $C = [c_1, c_2, \ldots, c_N]^T$ is defined as follows.

Definition 2 (Strongly Relevant, Weakly Relevant, and Irrelevant Feature, OSFS [15]): At time stamp t, the inflow of a feature F_t.

1) If $\forall S\subseteq F$-$\{F_t\}$ s.t. $P(C|S, F_t)\neq P(C|\ S)$, F_t is strongly relevant to C.
2) If $\exists S\subseteq F$-$\{F_t\}$ s.t. $P(C|S, F_t)\neq P(C|\ S)$, F_t is weakly relevant to C.
3) If $\forall S\subseteq F$-$\{F_t\}$ s.t. $P(C|S, F_t) = P(C|\ S)$, F_t is irrelevant to C.

Afterwards, the redundant features can be discarded by the Markov blanket defined as follows.

Definition 3 (Markov Blanket [15]): If X ($X\subset F$) is a Markov blanket for C, i.e., $M(C)$, they the following equalities hold:

$$\forall D \in F - X, P(C|X, D) = P(C|X). \tag{1}$$

TABLE I
Symbol Annotations

Symbol	Explanation
t	The time stamp $t\in\{1, 2, \ldots, \psi_t\}$.
F_t	The t-th vectors of streaming feature F.
$f_{n,t}$	The n-th element of F_t as $n\in\{1, 2, \ldots, N\}$.
F'_t	The t-th vectors of sparse streaming features F'.
$f'_{n,t}$	The n-th element of F'_t as $n\in\{1, 2, \ldots, N\}$.
$\hat{F}'_t$	Completed streaming feature of F'_t.
$p_{n,k}$	The n-th row and k-th column of P.
$q_{j,k}$	The j-th row and k-th column of Q.
L	The column number of buffer U.
h	The latent factor dimension of matrix P and Q.
λ	A regularization parameter of LFA.
η	The learning rate of LFA.
ζ_t	The missing data rate of sparse streaming features F'_t.
Λ_U	The known value set.
μ	The significance level.
$r_{\cdot *}$	The costs of three actions in states $\mathbb{C}$ and $\neg\mathbb{C}$ as $\cdot\in\{P, B, E\}$ and $*\in\{P, E\}$
$m^t_{\cdot *}$	The cardinal number when applying an action $a.$ to $[\mathbb{C}, \neg\mathbb{C}]$ as $\cdot\in\{P, B, E\}$ and $*\in\{P, E\}$.
α_t, β_t	The threshold of the relevance division at the timestamp t.

B. Three-Way Decision

The 3WD provides an effective tool for solving the uncertain problems. Its core idea is 'thinking in three' [41], i.e., dividing a set into three regions, which is formally defined as follows.

Definition 4 (Three-Way Decision [41]): Given a finite nonempty set W, it can be divided into tri-partitions denoted by $\pi = \{\text{POS}, \text{BND}, \text{NEG}\}$ by fulfilling the following conditions.

1) $\text{POS} \cap \text{BND} = \varnothing$, $\text{POS} \cap \text{NEG} = \varnothing$, $\text{BND} \cap \text{NEG} = \varnothing$.
2) $\text{POS} \cup BND \cup \text{NEG} = W$.

These three regions can be totally or partially ordered, which can be built via the evaluation models. With them, the three-way classification can be defined as follows.

Definition 5 (Three-Way Classification): Given an evaluation $e : W{\to}(R, \prec)$ on W as the e(x) denotes the evaluation value of x, then a pair of thresholds (β, α) $(\alpha, \beta \in R$ with $\beta{\prec}\alpha)$ can divide the three regions according to the following rules:

$$\text{POS}^{[\alpha,\cdot)}(e) = \left\{x \in W|e(x) \succeq \alpha\right\}$$

$$\text{BND}^{(\beta,\alpha)}(e) = \{x \in W|\beta \prec e(x) \prec \alpha\}$$

$$\text{NEG}^{(\cdot,\beta]}(e) = \left\{x \in W|e(x) \preceq \beta\right\}$$

where POS, BND, and NEG represent the positive, boundary, and negative region, respectively. Hence, 3WD can adopt different strategies depending on the characteristics of the three regions to improve the decision-making performance.

IV. Proposed Algorithm

A. Problem of 3WDO

In the context of financial market analysis, streaming stock data and economic indicators require rapid feature selection in order to adapt predictive models promptly to real-time

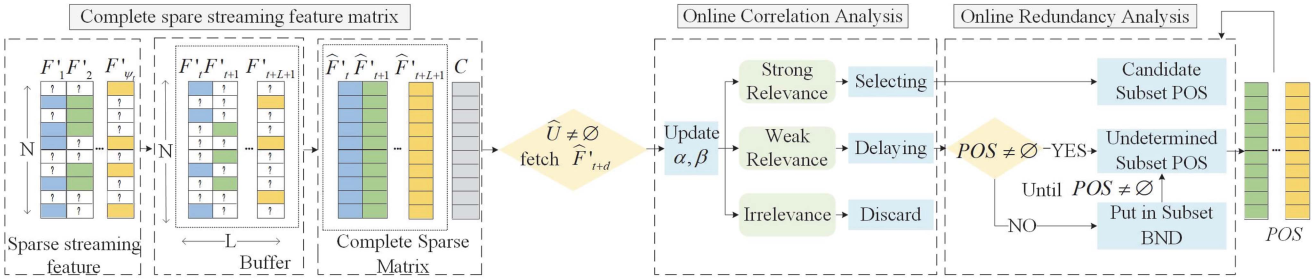

Fig. 1. Flowchart of 3WDO.

changes. However, some financial data might not be successfully obtained due to several reasons, including technical failures, network transmission interruptions, and incomplete records from the data sources themselves. Moreover, given the complex nature of the financial market where multiple factors are intertwined, it becomes challenging to precisely determine which factors exert a more dominant influence on target variables (such as stock returns, market index fluctuations, and so on), and which ones can be disregarded. This difficulty consequently introduces uncertainty into the feature selection process. Given $F' = \{F'_1, F'_2, \ldots, F'\psi_t\}$ to be sparse streaming features where the missing data rate of F'_t is calculated by $\zeta_t = 1-|\wedge_t|/N$; let $\wedge_U$ denote the known data set, $|\cdot|$ compute the cardinality of an enclosed set, and ζ be the total missing data rate of F'; then the challenge in OS^{2}FS is to build the minimum subset POS $\cup\{F'_t\}$. Therefore, 3WDO seeks for minimizing the decision risk of feature selection. To do so, the following major difficulties should be overcome.

1) How to use the LFA model to estimate the sparse streaming features with the minimum error.
2) How to implement the adaptive relevance analysis for addressing the uncertainty.
3) How to analyze the redundancy of different relevant features with the minimum decision risk.

B. Structure of 3WDO

The 3WDO model constitutes of three phases for high flexibility in OS^{2}FS: 1) phase I preprocesses the sparse streaming features; 2) phase II performs the online relevance analysis; and 3) phase III performs the online redundancy analysis. The processing flow of 3WDO is in Fig. 1.

1) Phase I Complete Spare Streaming Features: Suppose that sparse streaming features flow in $U^{N\times L}$ from time stamp t to $t+L-1$, defined as $U = \{F'_t, F'_{t+1}, \ldots, F'_{t+L-1}\}$, where N represents the sample size, and L is the column number. An LFA model [33], [34], [35], [36], [37], [38] is commonly applied to supplement missing data.

Definition 6 (LFA [34], [36]): The rank-h approximation $U = PQ^T$ is estimated relying on the known data set $\wedge_U$; i.e., the estimations to two latent factor matrices $P^{N\times h}$ and $Q^{L\times h}$ are obtained by training, and h stands for the latent factor dimension.

The LFA model tries to build a low-rank approximation $U = \{F'_t, F'_{t+1}, \ldots, F'_{t+L-1}\}$ for U. Commonly, to obtain matrix P and Q from $\wedge_U$, the minimum loss function is constructed by Euclidean distance between U and U be calculated as

$$\varepsilon(P, Q) = \frac{1}{2}\sum_{f'_{n,j}\in\Lambda_U}\left(f'_{n,j} - \sum_{k=1}^{h} p_{n,k}q_{j,k}\right)^2 + L \tag{2}$$

where $f'_{n,j} \in \wedge_U$ represents the known value, $p_{n,k}$ is the nth row and kth column of P, and $q_{j,k}$ is the jth row and kth column of $Q, j \in \{t, t+1, \ldots, t+L-1\}$ $\forall n \in \{1, 2\ldots, N\}$ $\forall\ k \in \{1, 2, \ldots, h\}$. L is a regularization scheme to avoid overfitting, so the L_2 scheme is used to improve the generalization ability [33]. And the calculation formula is as follows:

$$L = \frac{\lambda}{2}\left(\|P\|_F^2 + \|Q\|_F^2\right) \tag{3}$$

in which $\|\cdot\|_F$ computes the Frobenius norm, and λ denotes the regularization parameter.

Based on (2) and (3), predicting the complete streaming features according to the known value given by

$$\varepsilon = \sum_{f'_{n,j}\in\Lambda_U}\left(\frac{1}{2}\left(f'_{n,j} - \sum_{k=1}^{h} p_{n,k}q_{j,k}\right)^2 + \frac{\lambda}{2}\left(\sum_{k=1}^{h} p_{n,k}^2 + \sum_{k=1}^{h} q_{j,k}^2\right)\right). \tag{4}$$

Then, the loss function of the nth element $f'_{n,j}$ be compute by

$$\varepsilon_{n,j} = \frac{1}{2}\left(f'_{n,j} - \sum_{k=1}^{h} p_{n,k}q_{j,k}\right)^2 + \frac{\lambda}{2}\left(\sum_{k=1}^{h} p_{n,k}^2 + \sum_{k=1}^{h} q_{j,k}^2\right). \tag{5}$$

The stochastic gradient descent (SGD) is applied in solving the loss function [38], calculating the gradient of the loss function to the sum of parameters, and updating it in a descending direction

$$\begin{cases} p_{n,k} \leftarrow p_{n,k} - \eta\frac{\partial\varepsilon_{n,j}}{\partial p_{n,k}} \\ q_{j,k} \leftarrow q_{j,k} - \eta\frac{\partial\varepsilon_{n,j}}{\partial q_{j,k}} \end{cases}. \tag{6}$$

With (5) and (6), the partial derivative of the loss is obtained

$$\begin{aligned} p_{n,k} &\leftarrow p_{n,k} + \eta q_{j,k}\left(f'_{n,j} - \sum_{k=1}^{h} p_{n,k}q_{j,k}\right) - \lambda\eta p_{n,k} \\ q_{j,k} &\leftarrow q_{j,k} + \eta p_{n,k}\left(f'_{n,j} - \sum_{k=1}^{h} p_{n,k}q_{j,k}\right) - \lambda\eta q_{j,k}. \end{aligned} \tag{7}$$

TABLE II
COST MATRIX

Action	Cost Function	
	$\mathbb{C}$	$\neg\mathbb{C}$
a_P	r_{PP}	r_{PE}
a_B	r_{BP}	r_{BE}
a_E	r_{EP}	r_{EE}

TABLE III
CARDINAL NUMBER

Action	Cardinal Number	
	$\mathbb{C}$	$\neg\mathbb{C}$
a_P	m^t_{PP}	m^t_{PE}
a_B	m^t_{BP}	m^t_{BE}
a_E	m^t_{EP}	m^t_{EE}

where η is the learning rate, P and Q are trained to estimate the minimum errors on the known value, then $U = PQ^T$. The error between the estimated value and the actual data is $\text{err}_{n,j} = f'_{n,j} - \sum_{k=1}^{h} p_{m,k} q_{j,k}$, i.e.,

$$p_{n,k} \leftarrow p_{n,k} + \eta q_{j,k}\text{err}_{n,j} - \lambda\eta p_{n,k}$$
$$q_{j,k} \leftarrow q_{j,k} + \eta p_{n,k}\text{err}_{n,j} - \lambda\eta q_{j,k}. \quad (8)$$

2) Phase II Online Relevance Analysis: Existing OS^{2}FS models mainly concentrate on certain relationship between the labels and sparse features by two-way decision(2WD) [49], [50] (i.e., two-way relevance analysis). Expressly, a_P and a_E represent two actions of acceptance and rejection, which divide features into relevance (strong and weak relevance) or irrelevance. Under the null hypothesis of the conditional independence between C and F_t, with *ρ-value* returned by Fisher's Z-test [51] to measure conditional independence, a significance level μ judges the relevance. If $\rho \leq \mu$, feature F_t is divided into the relevant feature set $\mathbb{C}$; on the contrary, if $\rho > \mu$, then feature F_t is divided into the irrelevant feature set $\neg\mathbb{C}$. However, due to the uncertainty within the label-feature interactions, the conditional independence of the estimated sparse streaming features near the threshold μ may be misclassified. For this reason, the incorporated 3WD-based uncertain decision process (i.e., three-way relevance analysis) adds delayed action a_B to extend certain relevance and redundancy analysis. Delayed decision generates the rule for weakly relevant features. That is, moving the uncertain features to the boundary region as much as possible and waiting for more information, thus reducing the uncertainty risk

Different divisions of relevance analysis cause the corresponding cost. In practice, r_{PP}, r_{BP}, and r_{EP} denote the costs incurred for applying strategies a_P, a_B and a_E when a feature belongs to the relevant feature set $\mathbb{C}$. Analogously, r_{EE}, r_{BE}, and r_{PE} indicate the costs incurred for applying strategies a_P, a_B and a_E when a feature belongs to the irrelevant feature set $\neg\mathbb{C}$. As stated in [52], misclassification costs are more than the correct classification, i.e., $r_{\text{PP}} \leq r_{\text{BP}} \leq r_{\text{EP}}$ and $r_{\text{EE}} \leq r_{\text{BE}} \leq r_{\text{PE}}$. The cost matrix is shown in Table II.

The relevance between feature F_t and the class attribute C is $Dep(\mathbb{C}, F_t)$, and the irrelevance between feature F_t and the class attribute C is $\text{Dep}(\neg\mathbb{C}, F_t)$. Accordingly, the expected cost associated with applying strategies a_P, a_B and a_E be computed by

$$R(a_P|[F_t]) = r_{\text{PP}}\text{Dep}(\mathbb{C}, F_t) + r_{\text{PE}}\text{Dep}(\neg\mathbb{C}, F_t)$$
$$R(a_B|[F_t]) = r_{\text{BP}}\text{Dep}(\mathbb{C}, F_t) + r_{\text{BE}}\text{Dep}(\neg\mathbb{C}, F_t)$$
$$R(a_E|[F_t]) = r_{\text{EP}}\text{Dep}(\mathbb{C}, F_t) + r_{\text{EE}}\text{Dep}(\neg\mathbb{C}, F_t) \quad (9)$$

where $R(a_.|[F_t])$ $(\cdot \in \{P, B, E\})$ denotes the cost of feature F_t when taking action $a_.(\cdot \in \{P, B, E\})$.

According to the Bayesian decision rule, induced by the minimum-cost decision rules, the decision rules of relevance can be given by:

(P): If $R(a_P|[F_t]) \leq R(a_B|[F_t])$ and $R(a_P|[F_t]) \leq R(a_E|[F_t])$, the feature F_t is strong relevance, then decide $F_t \in$ POS;

(B): If $R(a_B|[F_t]) \leq R(a_P|[F_t])$ and $R(a_B|[F_t]) \leq R(a_E|[F_t])$, the feature F_t is weak relevance, then decide $F_t \in$ BND;

(E): If $R(a_E|[F_t]) \leq R(a_P|[F_t])$ and $R(a_E|[F_t]) \leq R(a_B|[F_t])$, the feature F_t is divided into irrelevance, then removing F_t.

In addition, the decision rules procedure suggests the following simplified feature relevance classification rules.

1) If $Dep(\mathbb{C}, F_t) \geq \alpha$, decide $F_t \in$POS, then the feature F_t is divided into strong relevance;
2) If $\beta < Dep(\mathbb{C}, F_t) < \alpha$, decide $F_t \in$BND, then the feature F_t is divided into weak relevance;
3) If $Dep(\mathbb{C}, F_t) \leq \beta$, the feature F_t is divided into irrelevance and discarded immediately.

And then, considering the cost of the three-way relevance analysis, the cardinal numbers of the above six conditions is defined in Table III, where actions a_P, a_B and a_E divide the features in set $\mathbb{C}$ into the total number of m^t_{PP}, m^t_{BP}, and m^t_{EP}, respectively. Similarly, actions a_P, a_B and a_E divide the features in set $\neg\mathbb{C}$ into the total number of m^t_{PE}, m^t_{BE}, and m^t_{EE}, respectively. Evidently, at the time stamp t, the decision cost of three-way relevance analysis is computed by

$$\text{COST}_t = r_{\text{EP}}m^t_{\text{EP}} + r_{\text{PE}}m^t_{\text{PE}} + r_{\text{BP}}m^t_{\text{BP}} + r_{\text{BE}}m^t_{\text{BE}} + r_{\text{PP}}m^t_{\text{PP}} + r_{\text{EE}}m^t_{\text{EE}}. \quad (10)$$

For convenience, extending from the assumptions of [52], the cost for the correct relevance classification is zero; i.e., $r_{PP} = r_{EE} = 0$. Based on (10), the cost function is given as

$$\text{COST}_t = r_{\text{EP}}m^t_{\text{EP}} + r_{\text{PE}}m^t_{\text{PE}} + r_{\text{BP}}m^t_{\text{BP}} + r_{\text{BE}}m^t_{\text{BE}} \quad (11)$$

where $\text{COST}^{mis}_t = r_{\text{EP}}m^t_{\text{EP}} + r_{\text{PE}}m^t_{\text{PE}}$ denotes the misclassification cost, and $\text{COST}^{del}_t = r_{\text{BP}}m^t_{\text{BP}} + r_{\text{BE}}m^t_{\text{BE}}$ denotes the delayed classification cost. The initial threshold values β and α of three-way relevance analysis are calculated according to [53]

$$\beta = \frac{(r_{\text{BE}} - r_{\text{EE}})}{(r_{\text{BE}} - r_{\text{EE}}) + (r_{\text{EP}} - r_{\text{BP}})}$$
$$\alpha = \frac{(r_{\text{PE}} - r_{\text{BE}})}{(r_{\text{PE}} - r_{\text{BE}}) + (r_{\text{BP}} - r_{\text{PP}})}. \quad (12)$$

To ensure the efficiency of three-way relevance analysis, finding an optimal threshold for feature F_t and generating the lowest risk cost in decision-making is vital. Thus, the problem of solving threshold parameters is an optimization problem as

$$\underbrace{\arg\min}_{(\beta_t, \alpha_t)}(\text{COST}_t). \quad (13)$$

Algorithm 1: Three-Way Relevance Threshold Algorithm

Input r_{PP}, r_{BP}, r_{EP}, r_{PE}, r_{BE}, r_{PE}, *minT*, *initT*, *delta*

Get a streaming feature $\mathrm{F_t}$ at time stamp t
$nowT = initT$
Initialize $COST_t$ according to (11)
Initialize β_t, α_t according to (12)
$\beta_t^{new} = \beta_t$, $\alpha_t^{new} = \alpha_t$
while $minT < nowT$
  Update β_t^{new}, α_t^{new}
  if $0 \leq \beta_t^{new} < \alpha_t^{new} \leq 1$
    Calculate $nowCOST_t$ according to (11)
  else
    Update β_t^{new}, α_t^{new} until $0 \leq \beta_t^{new} < \alpha_t^{new} \leq 1$
  if $COST_t - nowCOST_t > 0$ *or random.rand < exp(-res/(k*nowT))*
    $COST_t = nowCOST_t$
  $\beta_t = \beta_t^{new}$, $\alpha_t = \alpha_t^{new}$
  $nowT = nowT * delta$
Output β_t, α_t

Calculating the initial threshold according to (12) and automatically updating the threshold (β_t, α_t) by the simulated annealing algorithm, as shown in Algorithm 1. Hence, the initial threshold does not influence the result and can be initialized by itself. If the threshold reduces or keeps the COST_t unchanged, it is updated; otherwise, the threshold update is accepted with a certain probability, and the iterative process is repeated until it reaches the minimum decision cost. The time cost of this algorithm is $O(|initT|)$. In this algorithm, *initT* is the initial temperature. The 3WDO model sets $minT = 1$ and the temperature change as $nowT = nowT*delta$, where $delta = 0.95$ regulates the temperature [54].

3) Phase III Online Redundancy Analysis: The 3WDO model effectively reduces the decision cost by carrying on the uncertain redundant analysis according to the features' relevance, relying on the *ρ-value* returned by Fisher's Z-test to calculate the conditional dependence $Ind\ (C, F_t|X_F)$. If $\exists X_F \in M(C)_t$, s.t. $P(C|F_t, X_F) = P(C|X_F)$, concluding that F_t is a redundant feature, where $M(C)_t$ represents a Markov blanket of C. In what follows, based on the Markov blanket, Proposition 1 judges the redundance of features.

Proposition 1: For strongly relevant features, add them to POS and conduct a redundancy analysis between F_t and features in POS; if

$$\exists X_F \in \mathrm{POS}\ \text{s.t.}\ P(C|F_t, X_F) = P(C|X_F) \tag{14}$$

feature F_t is redundant and can be discarded. For weakly relevant feature F_t, when POS is nonempty set, if

$$\forall X_F \in \mathrm{POS}\ \text{s.t.}\ P(C|F_t, X_F) \neq P(C|X_F) \tag{15}$$

feature F_t is nonredundant and added to POS. Otherwise, when POS is an empty set, the weakly relevant features are put into BND until POS is a nonempty set, then the feature $F_i \in$POS and $F_j \in$BND are analyzed for redundancy.

Algorithm 2: 3WDO Algorithm

Initialize L, h, λ, η, ζ, maximum number of iterations *max*
1: **repeat**
2: get a sparse streaming feature F_t' at time stamp t
3: **if** $L \neq 0$
4:   $L = L - 1$, $U = U \cup F_t'$
5: /*complete the sparse streaming features matrix */
6: **while** $time = max$
7:   **for** $j = t$ *to* $t + L - 1$
8:     **for** $\forall f'_{n,j} \in \wedge_U$
9:       Update $p_{n,k}$, $q_{j,k}$ following (8)
10: $time = time - 1$, $U = PQ^T$, calculate β_t, α_t according to Algorithm 1
11: /* online sparse streaming features selection */
12: **for** $d = 1$ *to* L
13:   fetch F'_{t+d} form U
14:   **if** $Dep(\mathbb{C}, F'_{t+d}) \leq \beta_t$ *discard* F'_{t+d}
15:   **else if** $Dep(\mathbb{C}, F'_{t+d}) \geq \alpha_t$, $POS = POS \cup F'_{t+d}$
16:     discard redundant features F'_{t+d} following (14)
17:   **else**
18:     **if** $POS \neq \emptyset$
19:       add the features F'_{t+d} to the *POS* following (15)
20:     **else**
21:       $BND = BND \cup F'_{t+d}$
22:     until $POS \neq \emptyset$, add $F'_{t+d} \in BND$ to the *POS* following (15)
23:   **for** *each feature* $X_F \subseteq POS - F'_{t+d}$
24:     **if** $\exists b \subseteq POS - X_F$ *s. t. Ind* $(C, X_F|b)$
25:       $POS = POS - X_F$
26: $t = t + 1$, initialize $U = \emptyset$, L
27: **until** no features are available
Output *POS*

Nevertheless, irredundant features are added to POS, which may cause the redundancy of the already-selected features in POS. Consequently, Proposition 2 is proposed to check the redundancy of the already-selected ones in POS.

Proposition 2: At time point t, the feature F_t flow in POS, satisfying $M(F_t) \in |M(C)_t$. If

$$\begin{aligned} &\forall X_F \in M(C) \cup F_t, \exists \vartheta \subseteq M(C) \cup F_t - \{X_F\} \\ &\text{s.t.} \quad P(C|X_F. \vartheta) = P(C|\vartheta.\ ) \end{aligned} \tag{16}$$

then $\{X_F\}$ is redundant and be deleted.

C. Algorithm Design and Computational Complexity Analysis

Algorithm Design: Based on the above discussion, to improve the flexibility of the $\mathrm{OS^2FS}$ model, a new $\mathrm{OS^2FS}$ framework integrates with an improved 3WD, named 3WDO, as shown in Algorithm 2. In steps 2–4, get a new sparse streaming feature F_t' and put it in the buffer. When the buffer is filled, the LFA model completes the sparse streaming features as given in steps 6–10. Then, updating the thresholds β_t and α_t in terms of *Algorithm* 1 and carrying out uncertain relevance and redundancy analysis. If $Dep(\mathbb{C}, F'_{t+d}) \le \beta_t$, F'_{t+d} is the irrelevant feature and will be discarded immediately in step 14. If $Dep(\mathbb{C}, F'_{t+d}) \ge \alpha_t$, F'_{t+d} is the strongly relevant feature and be added to POS in step 15. Furthermore, check whether F'_{t+d} is a redundant feature in step 16. If $\beta_t < Dep(\mathbb{C}, F'_{t+d}) < \alpha_t$, F'_{t+d} is the weakly relevant feature. When POS is nonempty, the irredundant feature is added to POS in step 19. On the contrary, add F'_{t+d} into BND until POS is nonempty, and then check whether there are features in BND and POS are irredundant and redundant features in steps 22 and 23–25.

The Computational Complexity Analysis: The 3WDO model takes a little more time than the traditional OS^2FS models in Phase I; it takes time to handle the missing data. Given ψ_t sparse streaming features with missing data rate ζ_t have N instances at timestamp t. The time complexity of Phase I is $O(N \times \psi_t \times (1\text{-}\zeta_t) \times h)$. This model compares $Dep(\mathbb{C}, F'_{t+d})$ with β_t and α_t in Phase II, and the time complexity of this stage is $O(L)$. In Phase III, if $Dep(\mathbb{C}, F'_{t+d}) > \beta_t$ and POS$\neq \emptyset$, conducting redundancy analysis between feature F'_{t+d} with the already-selected features in POS. The worst time is $O(|\text{POS}|)$. Otherwise, the w*eakly relevant* features are put into BND until POS$\neq \emptyset$, *comparing* features in BND with already-selected features. The worst time complexity of this process is $O(|\text{POS}| \times |\text{BND}|/2)$. The space complexity of 3WDO is $O(\psi_t)$.

D. Theoretical Analyses Regarding 3WDO

1) Theoretical Analysis of Sparse Feature Completion: This section provides a theoretical examination of the convergence in Phase I, focusing on transforming the sparse streaming feature matrix U into the complete streaming features U. Among them, the missing value in U can be random, nonrandom or complex missing, etc. (This *theoretical analysis* is proved in the supplementary file).

2) Theoretical Analyses Regarding 3WDO's Cost: In the following discussions, an interpretable perspective is provided to demonstrate the three-way relevance analysis performs better than the two-way.

1) *The Decision Cost of Relevance Analysis:* With the cost matrix and the cardinal number of three-way relevance analysis, the cost matrix can be given in Table IV. In addition, the cardinal number of these four conditions is constructed in Table V. Correspondingly, at timestamp t, the decision cost of two-way relevance analysis is computed as

$$\mathrm{COST}_t^{(2)} = r_{\mathrm{PP}}^{(2)} * m_{\mathrm{PP}}^{t,(2)} + r_{\mathrm{PE}}^{(2)} * m_{\mathrm{PE}}^{t,(2)} + r_{\mathrm{EP}}^{(2)} * m_{\mathrm{EP}}^{t,(2)} + r_{\mathrm{EE}}^{(2)} * m_{\mathrm{EE}}^{t,(2)}. \tag{17}$$

Similarly, $r_{\mathrm{PP}}^{(2)} = r_{\mathrm{EE}}^{(2)} = 0$, (17) is simplified to

$$\mathrm{COST}_t^{(2)} = r_{\mathrm{PE}}^{(2)} * m_{\mathrm{PE}}^{t,(2)} + r_{\mathrm{EP}}^{(2)} * m_{\mathrm{EP}}^{t,(2)} \tag{18}$$

Three-way relevance analysis moves misclassification relevant features into boundary region, and the correct classification features remain unchanged, which is formulated as:

$$m_{\mathrm{PE}}^{t,(2)} = m_{\mathrm{PE}}^{t} + m_{\mathrm{BE}}^{t}$$
$$m_{\mathrm{EP}}^{t,(2)} = m_{\mathrm{EP}}^{t} + m_{\mathrm{BP}}^{t}$$
$$m_{\mathrm{PP}}^{t,(2)} = m_{\mathrm{PP}}^{t}$$

TABLE IV
COST MATRIX OF TWO-WAY

Action	Cost Function	
	$\mathbb{C}$	$\neg\mathbb{C}$
a_P	$r_{PP}^{(2)}$	$r_{PE}^{(2)}$
a_E	$r_{EP}^{(2)}$	$r_{EE}^{(2)}$

TABLE V
CARDINAL NUMBER OF TWO-WAY

Action	Cardinal Number	
	$\mathbb{C}$	$\neg\mathbb{C}$
a_P	$m_{PP}^{t,(2)}$	$m_{PE}^{t,(2)}$
a_E	$m_{EP}^{t,(2)}$	$m_{EE}^{t,(2)}$

TABLE VI
CONFUSION MATRIX

Confusion matrix		Predicted Class	
		Class=positive	Class=negative
Actual Class	positive	PP	PE
	negative	EP	EE

$$m_{\mathrm{EE}}^{t,(2)} = m_{\mathrm{EE}}^{t}. \tag{19}$$

Proposition 3: The decision cost of the three-way relevance analysis is less than that of the two-way, i.e., $\mathrm{COST}_t < \mathrm{COST}_t^{(2)}$.

Proof of Proposition 3: Based on (11) and (19), the following inferences are obtained:

$$\begin{aligned}\mathrm{COST}_t - \mathrm{COST}_t^{(2)} &= \left(r_{\mathrm{EP}} m_{\mathrm{EP}}^t + r_{\mathrm{PE}} m_{\mathrm{PE}}^t + r_{\mathrm{BP}} m_{\mathrm{BP}}^t + r_{\mathrm{BE}} m_{\mathrm{BE}}^t\right)\\ &\quad - \left(r_{\mathrm{PE}} * m_{\mathrm{PE}}^{t,(2)} + r_{\mathrm{EP}} * m_{\mathrm{EP}}^{t,(2)}\right)\\ &= r_{\mathrm{EP}} * \left(m_{\mathrm{EP}}^t - m_{\mathrm{EP}}^{t,(2)}\right) + r_{\mathrm{PE}} * \left(m_{\mathrm{PE}}^t - m_{\mathrm{PE}}^{t,(2)}\right)\\ &\quad + r_{\mathrm{BP}} m_{\mathrm{BP}}^t + r_{\mathrm{BE}} m_{\mathrm{BE}}^t\\ &= m_{\mathrm{BP}}^t * (r_{\mathrm{BP}} - r_{\mathrm{EP}}) + m_{\mathrm{BE}}^t * (r_{\mathrm{BE}} - r_{\mathrm{PE}})\end{aligned} \tag{20}$$

Since $r_{\mathrm{BP}} < r_{\mathrm{EP}}$ and $r_{\mathrm{BE}} < r_{\mathrm{PE}}$, then $\mathrm{COST}_t\text{-}\mathrm{COST}_t^{(2)} < 0$, i.e., $\mathrm{COST}_t < \mathrm{COST}_t^{(2)}$. The cost of the three-way relevance analysis is less than that of the two-way, which effectively reduces decision risks. Therefore, Proposition 3 holds. ■

2) *The Comparative Analysis on Confusion Matrix:* A confusion matrix is adopted to evaluate the effectiveness of a three-way relevance analysis. The confusion matrix is given in Table VI, where PP stands for prediction in the positive domain and actual position in the positive domain, while EE indicates prediction in the negative domain, actual position in the negative domain, and so on.

Using the above indicators, the three-way relevance analysis and two-way comparison results are as follows: $Precision^{(2)} < Precision$, $Recall^{(2)} = Recall$, $Accuracy^{(2)} = Accuracy$, $F_2\ 1 < F_1$.

Proof of Proposition 4: According to the confusion matrix, *Precision*, *Recall*, *Accuracy*, and F_1 score evaluate classification performance.

1) Based on $m_{\mathrm{PP}}^{t,(2)} = m_{\mathrm{PP}}^t$ and $m_{\mathrm{PE}}^{t,(2)} > m_{\mathrm{PE}}^t$, we have: $\mathrm{Precision}^{(2)} = \frac{m_{\mathrm{PP}}^{t,(2)}}{m_{\mathrm{PP}}^{t,(2)} + m_{\mathrm{PE}}^{t,(2)}} < \frac{m_{\mathrm{PP}}^t}{m_{\mathrm{PP}}^t + m_{\mathrm{PE}}^t} = \mathrm{Precision}$;
2) In terms of $m_{\mathrm{PP}}^{t,(2)} = m_{\mathrm{PP}}^t$, $m_{\mathrm{EP}}^{t,(2)} = m_{\mathrm{EP}}^t + m_{\mathrm{BP}}^t$, we have: $\mathrm{Recall}^{(2)} = \frac{m_{\mathrm{PP}}^{t,(2)}}{m_{\mathrm{PP}}^{t,(2)} + m_{\mathrm{EP}}^{t,(2)}} = \frac{m_{\mathrm{PP}}^t}{m_{\mathrm{PP}}^t + m_{\mathrm{BP}}^t + m_{\mathrm{EP}}^t} = \mathrm{Recall}$;

3) On the basis of $m_{\mathrm{PP}}^{t,(2)} = m_{\mathrm{PP}}^{t}$, $m_{\mathrm{EE}}^{t,(2)} = m_{\mathrm{EE}}^{t}$, the cardinality of all samples is N, we have: $\text{Accuracy}^{(2)} = \frac{m_{\mathrm{PP}}^{t,(2)}+m_{\mathrm{EE}}^{t,(2)}}{N} = \frac{m_{\mathrm{PP}}^{t}+m_{\mathrm{EE}}^{t}}{N} = \text{Accuracy}$;
4) According to $m_{\mathrm{PP}}^{t,(2)} = m_{\mathrm{PP}}^{t}$, $m^{t},(2)\ EP = m_{\mathrm{EP}}^{t} + m_{\mathrm{BP}}^{t}$, $m_{\mathrm{PE}}^{t,(2)} > m_{\mathrm{PE}}^{t}$, we have

$$F_1^{(2)} = \frac{2m_{\mathrm{PP}}^{t,(2)}}{2m_{\mathrm{PP}}^{t,(2)} + m_{\mathrm{EP}}^{t,(2)} + m_{\mathrm{PE}}^{t,(2)}} < \frac{2m_{\mathrm{PP}}^{t}}{2m_{\mathrm{PP}}^{t} + m_{\mathrm{BP}}^{t} + m_{\mathrm{EP}}^{t} + m_{\mathrm{PE}}^{t}} = F_1. \quad (21)$$

Therefore, three-way relevance analysis performs better than two-way under indicators *Precision* and F_1 score and performance similarly under indicators *Recall* and *Accuracy*, *Proposition* 4 stands. ■

V. Experiments and Results

In the subsequent experiments, the following problems are mainly solved.

1) *R.Q. 1.* Does the 3WDO model significantly outperform the other OSFS and OS^{2}FS models with the increased missing rate?
2) *R.Q. 2.* Does the 3WDO model of three-way relevance analysis perform better than state-of-the-art OSFS and OS^{2}FS models?
3) *R.Q. 3.* How effective is the completion of the spare streaming features of the 3WDO algorithm?
4) *R.Q. 4.* Does the 3WDO model perform better than the other OSFS and OS^{2}FS models with nonrandomly distributed missing data?
5) *R.Q. 5.* How does the parameter influence the performance of the 3WDO model?

A. General Settings

Datasets: In this section, we use twelve real-world data sets from DNA microarray data, NIPS 2003 data, and studies in [55], as shown in Table VII.

Baselines: We choose seven state-of-the-art OSFS and OS^{2}FS models to validate the performance of the 3WDO model, including Fast-OSFS, SAOLA, SFS-FI, OSSFS-DD, RHOFS, ANOHFS and LOSSA. Besides, three basic classifiers, support vector machine (SVM), k-nearest neighbors (KNN), and random forest (RF), are adopted to evaluate the validity of feature selection. Tables VIII and IX summarize the parameters set to the values of these algorithms and classifiers. These algorithms are conducted on MATLAB. We perform five-fold cross-validation in these experiments, where 4/5 data is the training set, and 1/5 data is the test set. Repeating each data set ten times and reporting the result of predictive accuracy. All experiments are implemented on our personal computer (Intel i7 2.40-GHz CPU, RAM 16GB).

The Computational Complexity Analysis: The best time complexity of Fast-OSFS is $O(|SF|k^{|PSO|})$ (where $|SF|$ denotes the relevant feature count, $k^{|PSO|}$ represents the size of all subsets in POS is less than or equal to k) if the second part is not applied. The worst time complexity of the second part is $O(|SF||PSO|k^{|PSO|*})$ (where $k^{|PSO|*}$ denotes the size of all subsets in POS does not exceed k). The time complexity of SAOLA is $O(\psi_t|\mathrm{POS}|)$. The worst time complexity of SFS-FI is $O(\psi_t{}^2)$. The time complexity of uncertainty analysis in OSSFS-DD is $O(|\mathrm{BND}|^2/2)$, and removing redundant features from the candidate feature subset is $O(|\mathrm{POS}|^2/2)$. The worst time complexity of RHOFS is $O(mu^2n'c')$ (where mu denotes the size of conditional features, n', c' represent the number of objects and classes, respectively). In the best case, the ANOHFS model's time complexity is $O(|m^*|^2|Subtr| \lg |Subtr|)$. In the worst case, F_t' needs an online redundancy analysis, requiring traversing all features in POS, and the time complexity is $O(|\mathrm{POS}||m^*|^2|Subtr| \lg |Subtr|)$ (where $|Subtr|$ is the number of instances in the subtree instance set, $|m^*|$ is the number of instances in U). The time complexity of the first phase of LOSSA is $O(N\times\psi_t\times(1\text{-}\zeta_t)\times h)$. This experiment combines it with Fast-OSFS, and the time complexity of the second stage is consistent with Fast-OSFS. The space complexity of all the above algorithms is $O(\psi_t)$.

Experimental Designs: The 3WDO model compares with the algorithms above on sparse streaming features. According to the parametric experiment, the default settings are set as $init = 10$, $\lambda = 0.01$, $h = 10$, and $L = 5$. The initial threshold is calculated according to the cost matrix, and then the optimal threshold is found according to algorithm 1. The cost matrix is only used as a starting value to calculate the threshold so that the cost matrix can be set to any value, such as r_{PP}=0,

TABLE VII
Details of Selected Datasets

Mark	Dataset	#(Features)	#(Instances)	#(C lass)
D1	Madelon	501	2600	6
D2	HAPT	561	10929	12
D3	Marti1	1025	500	2
D4	USPS	1500	242	2
D5	Colon	2001	62	2
D6	SRBCT	2309	83	4
D7	Lung	3313	83	5
D8	Prostate	6033	102	2
D9	Leukemia	7071	72	2
D10	ALLAML	7130	72	2
D11	CLL_SUB_111	11340	111	3
D12	Lungcancer	12534	181	2

TABLE VIII
Algorithm Parameters Adopted in the Experiments

Mark	Algorithm	Parameter
M1	3WDO	*Z test*, *Alpha is* 0.05.
M2	Fast-OSFS [15]	*Z test*, *Alpha is* 0.05 (TPAMI, 2013).
M3	SAOLA [16]	*Z test*, *Alpha is* 0.05 (TKDD, 2016).
M4	SFS-FI [25]	*Z test*, *Alpha is* 0.05, γ=0.05 (TNNLS, 2021).
M5	OSSFS-DD [26]	*Z test*, *Alpha is* 0.05 (TKDD, 2022).
M6	RHOFS [29]	None (TPDS, 2023).
M7	ANOHFS [30]	None (ASC, 2024).
M8	LOSSA [19]	*Z test*, *Alpha is* 0.05, λ=0.01, B_s=10 (TSMC, 2022).

TABLE IX
All the Classifier Parameters Used in the Experiments

Classifier	Parameter
KNN	*The neighbors* are 3.
SVM	*Default parameters values.*
RF	*6 decision trees.*

TABLE X
RANK SUM OF THE WILCOXON SIGNED-RANKS WHEN ζ FROM 0.1 TO 0.9 ON ALL THE DATASETS

ζ	Fast-OSFS#		SAOLA		SFS-FI		OSSFS-DD		RHOFS		ANOHFS		LOSSA	
	R^+	R^-	R^+	R^-	R^+	R^-	R^+	R^-	R^+	R^-	R^+	R^-	R^+	R^-
0.1	78	0	73	5	78	0	78	0	78	0	78	0	73	5
0.2	72	6	70	8	78	0	78	0	74	4	77	1	71	7
0.3	78	0	78	0	78	0	78	0	76	2	78	0	71	7
0.4	60	6	75	3	78	0	77	1	78	0	78	0	68	10
0.5	64	4	77	1	78	0	78	0	69	9	78	0	71	7
0.6	53	13	76	2	78	0	78	0	68	10	78	0	72	6
0.7	43	23	66	12	78	0	78	0	77	1	78	0	71	7
0.8	56	10	70	8	77	1	78	0	61	17	78	0	72	6
0.9	51	15	65	13	75	3	78	0	66	12	77	1	74	4

*If min{R^+, R^-}>18, the null hypothesis will be taken; #after the missing ratio of D2 exceeds 0.4, Fast-OSFS could not obtain a round of results for two weeks.

$r_{BP} = 1$, $r_{EP} = 10$, $r_{PE} = 10$, $r_{BE} = 1$, $r_{EE} = 0$ [41]. To further check the performance of the different algorithms, conduct the Friedman test at a 95% significance level under the null hypothesis. Moreover, paired Wilcoxon signed-ranks test at 0.1 significance level to analyze whether significant differences exist between the 3WDO model and others. The *ρ-value* represents the 3WDO model outperforming all these competing algorithms.

B. 3WDO Versus OSFS and OS2FS (R.Q. 1, R.Q. 2)

1) Accuracy Analysis With Missing Data Rates From 0.1 to 0.9: To test the effectiveness of the 3WDO model, we compare our algorithm with seven representative OSFS and OS^2FS models with missing data rates from 0.1 to 0.9 on twelve datasets. They are Fast-OSFS, SAOLA, SFS-FI, OSSFS-DD, RHOFS, ANOHFS, and LOSSA. Except for LOSSA, these algorithms focus on complete streaming feature scenarios. In this article, zero-filling primarily deals with the massive missing data for Fast-OSFS, SAOLA, SFS-FI, OSSFS-DD, RHOFS, and ANOHFS. We highlight the results where these algorithms have higher accuracy than other algorithms. The average classification accuracy of the 3WDO model with KNN, SVM, and RF on all the datasets is presented in Table S(I) in the supplementary file. The last row is the average accuracy of the 3WDO model on all data sets, which shows the accuracy change trend of the missing rate from 0.10 to 0.90. Table X records the results of 3WDO and other algorithms compared one by one by the Wilcoxon signed-ranks test. Fig. S1 in the supplementary file shows the average accuracy of all algorithms on D1-12, in which the Fast-OSFS algorithm has not produced a round of results for two weeks after the missing value ratio of D2 exceeds 0.4.

The Average Accuracy of the 3WDO Model on All the Datasets: The accuracy decreased slightly with the increase in missing data rate. However, the accuracy of some data sets floats a little with the rise of the missing data rate. For example, the accuracy of D3 has remained around 0.86.

The Wilcoxon Signed-Ranks Test: The Wilcoxon signed-ranks test for the average accuracy of the 3WDO compared with other methods is recorded in Table X. The results show that our algorithm is better than other algorithms on most data sets when the missing data rate increases from 0.1 to 0.9.

The Average Accuracy of All Algorithms on D1-12: From the and Fig. S1 in the supplementary file, we have the following findings.

1) The accuracy of most algorithms decreases with the increased missing data rate. Compared with other algorithms, the accuracy of the 3WDO model is superior to its rivals when the missing data rate increases from 0.1 to 0.9 on most datasets. The sparse streaming features completed by zero-filling are imprecise, making Fast-OSFS, SAOLA, SFS-FI, OSSFS-DD, RHOFS, and ANOHFS performance inferior to the 3WDO model.
2) The average accuracy of LOSSA is higher than Fast-OSFS, SAOLA, SFS-FI, OSSFS-DD, RHOFS, and ANOHFS when the missing data rate is from 0.1 to 0.6 on most datasets. The reason is that there are few known data, and the error of the incomplete streaming features pre-estimated by the LFA model increases. In certain relevance analysis, essential features may be wrongly classified as irrelevant and directly deleted. The 3WDO model efficiently solves this problem by reasonably dividing features and reducing the decision risk. It is ensured the higher accuracy of the 3WDO model.

In sum, the 3WDO model pre-estimates the missing data through the LFA model, which improves the accuracy of the traditional OS^2FS model.

2) Compare M1 with M8: M1 and M8 use the LFA model to complete missing values. To better compare M1 with M8, the situations of selected features by M1 and M8 versions are shown in Table XI. Then, we further compare the influence of features selected by M1 and M8 on the accuracy, and the conditional independence of different features selected by M1 and M8 on class attributes is shown in Table XII.

The Situations of Selected Features by M1 and M8 Versions: M8 selects more features than M1. This is mainly because certain relevance and redundancy analyses choose redundant features when dealing with sparse streaming features. With the increase in the missing rate, M1 and M8 select fewer and fewer identical features. When the missing rate increases, the completion data error of M8 increases the misclassifying features probability.

The M1 and M8 select the conditional independence of different features. According to Table XII, the conditional independence of M1 is lower than that of M8 on most data sets with the missing rate from 0.1 to 0.9, which proves that the features selected by M1 are more related to class attributes than M8. The result is consistent with the conclusion of Proposition 3. The reason is that the decision cost of the three-way relevance analysis is less than that of the two-way. Accordingly, the 3WDO obtains high accuracy.

TABLE XI
SITUATIONS OF SELECTED FEATURES BY M1 AND M8 VERSIONS

Dataset	0.1			0.2			0.3			0.4			0.5			0.6			0.7			0.8			0.9		
	M1	M8	Δ	M1	M8	Δ	M1	M8	Δ	M1	M8	Δ	M1	M8	Δ	M1	M8	Δ	M1	M8	Δ	M1	M8	Δ	M1	M8	Δ
D1	12.0	15.0	4.0	12.3	15.0	6.0	12.0	13.0	3.0	12.0	14.0	3.0	10.7	12.0	1.0	12.3	13.3	2.0	13.7	15.0	2.0	11.0	13.0	1.0	10.7	11.0	0
D2	11.0	8.0	4.0	14.0	13.0	6.0	14.0	15.0	7.0	17.1	25.0	5.0	24.0	46.0	8.0	42.0	61.0	8.0	54.1	88.0	13.0	56.0	144	15.0	56.6	158	20.0
D3	3.7	4.0	0	3.7	6.0	0	5.0	8.0	0	6.0	7.0	0	4.3	7.0	0	6.3	7.0	0	9.7	11.0	0	8.0	12.0	0	7.3	9.0	0
D4	9.0	11.0	4.0	9.0	11.0	4.0	12.0	13.0	4.0	13.0	14.0	5.0	11.0	17.0	5.0	17.0	19.0	7.0	13.0	16.0	3.0	10.0	13.0	3.0	10.0	11.0	0
D5	3.0	4.0	1.0	3.0	4.0	1.0	3.0	5.0	0	3.7	4.0	0	3.0	5.0	0	3.7	5.0	0	4.0	5.0	0	3.7	5.0	0	4.3	5.0	0
D6	5.0	7.0	1.0	7.0	8.0	2.0	5.7	8.0	1.0	5.7	8.0	0	7.3	9.0	0	5.7	8.0	0	6.0	8.0	0	5.0	7.0	0	6.0	7.0	0
D7	7.0	14.0	7.0	10.0	17.0	7.0	12.3	20.0	6.0	12.0	21.0	2.0	16.0	19.0	2.0	14.7	16.0	1.0	12.3	15.0	0	11.0	12.0	0	7.3	8.0	0
D8	4.3	5.0	1.0	5.0	8.0	1.0	5.0	8.0	1.0	6.0	8.0	1.0	6.0	8.0	1.0	6.7	7.0	0	7.0	9.0	0	7.7	8.0	0	7.3	8.0	0
D9	6.0	9.0	0	7.7	9.0	0	8.3	10.0	1.0	9.0	10.0	1.0	9.0	10.0	0	8.0	9.0	1.0	9.0	10.0	0	7.0	9.0	0	6.3	9.0	0
D10	6.7	8.0	1.0	7.0	8.0	2.0	7.0	8.0	0	7.0	8.0	0	7.0	8.0	0	7.0	8.0	0	8.0	9.0	2.0	7.0	8.0	0	6.7	7.0	0
D11	10.0	10.0	1.0	8.2	10.0	0	7.8	8.7	1.0	9.3	9.0	0	9.0	6.0	0	6.0	10.0	0	8.8	8.0	0	8.0	9.0	0	7.6	11.0	0
D12	9.0	10.0	2	12.0	16.0	4.0	14.0	17.0	5.0	16.0	17.0	5.0	18.0	20.0	4.0	19.0	21.0	1.0	17.0	23.0	0	16.0	21.0	1.0	11.3	16.0	0

Δ The number of same features.

TABLE XII
CONDITIONAL INDEPENDENCE OF DIFFERENT FEATURES SELECTED BY M1 AND M8

Dataset	0.1		0.2		0.3		0.4		0.5		0.6		0.7		0.8		0.9	
	M1	M8	M1	M8	M1	M8	M1	M8	M1	M8	M1	M8	M1	M8	M1	M8	M1	M8
D1	0.0498	0.0675	0.0508	0.0760	0.0415	0.0682	0.0378	0.0597	0.0307	0.0488	0.0212	0.0229	0.0263	0.0446	0.0197	0.0284	0.0212	0.0231
D2	0.4730	0.5129	0.4117	0.4918	0.4267	0.4286	0.3670	0.5138	0.3243	0.4680	0.2738	0.3576	0.2386	0.2926	0.1595	0.2004	0.1052	0.1313
D3	0.0291	0.0354	0.0203	0.0286	0.0141	0.0197	0.0185	0.0282	0.0142	0.0170	0.0277	0.0322	0.0128	0.0263	0.0206	0.0249	0.0198	0.0348
D4	0.0993	0.1026	0.0997	0.1025	0.0807	0.0821	0.0612	0.0624	0.0694	0.0760	0.0295	0.0578	0.0429	0.0589	0.0388	0.0464	0.0074	0.0238
D5	0.0894	0.1128	0.0889	0.1089	0.1458	0.1586	0.1097	0.1159	0.0639	0.1918	0.0678	0.1371	0.0688	0.0982	0.0473	0.0782	0.0542	0.1185
D6	0.0513	0.1257	0.0621	0.1286	0.0817	0.0968	0.0801	0.1363	0.0863	0.1134	0.0866	0.1422	0.0733	0.0924	0.0610	0.0731	0.0740	0.0762
D7	0⁑	0.0867	0.0831	0.1155	0.1443	0.1655	0.0975	0.1171	0.0963	0.1238	0.0923	0.0996	0.0403	0.1155	0.0295	0.0549	0.0339	0.0433
D8	0.1261	0.1883	0.0998	0.1254	0.0613	0.0919	0.1216	0.1308	0.0955	0.1030	0.0685	0.1886	0.0485	0.1263	0.0749	0.1518	0.0641	0.1204
D9	0.0969	0.1073	0.0289	0.1383	0.0944	0.1189	0.0823	0.1083	0.0775	0.1478	0.0732	0.1400	0.0776	0.1273	0.0726	0.1390	0.0636	0.1087
D10	0.0519	0.0798	0.0559	0.0849	0.1077	0.1122	0.1325	0.1520	0.1235	0.1624	0.1014	0.1300	0.1175	0.1331	0.0549	0.0925	0.0778	0.0803
D11	0.4146	0.3987	0.3435	0.2944	0.2833	0.3279	0.2426	0.2798	0.1647	0.2414	0.1045	0.1628	0.1108	0.1871	0.0757	0.1156	0.0712	0.0783
D12	0.0983	0.1025	0.1048	0.1187	0.1488	0.1548	0.0817	0.1690	0.1144	0.1480	0.1126	0.1556	0.1419	0.1554	0.1276	0.1370	0.1143	0.1250

⁑ The features selected by M1 are all in M8.

TABLE XIII
USING THE SELECTED FEATURES TO TRAIN THREE CLASSIFIERS AND THE TESTING ITS ACCURACY (%), $\zeta = 0.1$

Models/Datasets		D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	Average
KNN	M1	**58.10**$_{\pm0.50}$	**77.56**$_{\pm0.36}$	**86.20**$_{\pm0.56}$	88.46$_{\pm0.43}$	84.97$_{\pm0.20}$	85.56$_{\pm3.24}$	**88.23**$_{\pm1.40}$	**94.54**$_{\pm1.71}$	**98.90**$_{\pm0.77}$	**98.50**$_{\pm1.38}$	61.19$_{\pm2.73}$	**98.89**$_{\pm0.31}$	**85.09**$_{\pm1.13}$
	M2	56.40$_{\pm0.14}$	73.55$_{\pm0.08}$	84.90$_{\pm0.67}$	86.80$_{\pm0.27}$	77.94$_{\pm2.46}$	81.93$_{\pm0.70}$	81.13$_{\pm0.85}$	92.04$_{\pm1.68}$	98.33$_{\pm0.05}$	96.10$_{\pm0.05}$	63.34$_{\pm3.60}$	98.07$_{\pm0.33}$	82.54$_{\pm0.91}$
	M3	54.47$_{\pm0.45}$	59.52$_{\pm0.34}$	77.92$_{\pm0.61}$	82.61$_{\pm0.49}$	80.96$_{\pm2.89}$	82.88$_{\pm2.00}$	82.06$_{\pm0.79}$	91.88$_{\pm0.55}$	98.74$_{\pm0.01}$	97.63$_{\pm0.71}$	55.69$_{\pm3.49}$	98.45$_{\pm0.01}$	80.23$_{\pm1.03}$
	M4	49.76$_{\pm0.22}$	50.10$_{\pm0.37}$	77.13$_{\pm0.31}$	74.41$_{\pm0.59}$	64.51$_{\pm0.68}$	57.49$_{\pm0.79}$	63.70$_{\pm1.03}$	73.51$_{\pm3.46}$	34.65$_{\pm0.20}$	67.01$_{\pm4.14}$	51.89$_{\pm3.07}$	87.69$_{\pm1.80}$	62.65$_{\pm1.39}$
	M5	49.69$_{\pm0.42}$	47.08$_{\pm0.16}$	85.14$_{\pm0.49}$	81.65$_{\pm0.61}$	85.37$_{\pm1.49}$	85.42$_{\pm1.70}$	85.66$_{\pm1.65}$	93.17$_{\pm0.38}$	98.62$_{\pm0.92}$	98.49$_{\pm0.44}$	54.43$_{\pm2.61}$	97.57$_{\pm0.54}$	77.20$_{\pm1.06}$
	M6	52.22$_{\pm0.24}$	65.08$_{\pm3.33}$	85.22$_{\pm0.60}$	**88.65**$_{\pm0.34}$	85.62$_{\pm2.05}$	**92.26**$_{\pm3.08}$	79.16$_{\pm3.24}$	90.67$_{\pm1.40}$	97.21$_{\pm1.12}$	93.26$_{\pm0.99}$	61.23$_{\pm3.14}$	96.07$_{\pm1.26}$	82.22$_{\pm1.73}$
	M7	55.79$_{\pm0.79}$	77.28$_{\pm0.15}$	85.44$_{\pm0.51}$	78.97$_{\pm0.75}$	79.47$_{\pm2.33}$	83.55$_{\pm1.50}$	80.34$_{\pm1.27}$	91.20$_{\pm1.32}$	97.23$_{\pm0.07}$	96.95$_{\pm1.12}$	61.63$_{\pm4.64}$	98.28$_{\pm0.18}$	82.18$_{\pm1.22}$
	M8	55.23$_{\pm0.67}$	72.40$_{\pm0.18}$	86.06$_{\pm0.75}$	86.23$_{\pm0.20}$	**88.28**$_{\pm1.78}$	83.24$_{\pm0.04}$	83.50$_{\pm1.07}$	91.57$_{\pm0.95}$	98.60$_{\pm0.05}$	98.06$_{\pm0.73}$	**64.34**$_{\pm3.64}$	98.34$_{\pm0.01}$	83.82$_{\pm0.84}$
SVM	M1	**59.51**$_{\pm0.52}$	**72.70**$_{\pm0.06}$	**88.20**$_{\pm0.01}$	**84.93**$_{\pm0.22}$	**85.01**$_{\pm2.38}$	83.49$_{\pm1.29}$	**87.05**$_{\pm1.97}$	**89.72**$_{\pm2.59}$	**98.60**$_{\pm1.73}$	85.42$_{\pm4.49}$	69.58$_{\pm4.09}$	97.24$_{\pm0.32}$	**83.45**$_{\pm1.64}$
	M2	59.03$_{\pm0.39}$	71.23$_{\pm0.07}$	71.30$_{\pm4.19}$	80.00$_{\pm0.21}$	78.77$_{\pm1.68}$	77.24$_{\pm2.66}$	80.35$_{\pm1.71}$	88.30$_{\pm1.00}$	94.59$_{\pm0.82}$	87.17$_{\pm2.56}$	53.69$_{\pm8.27}$	97.19$_{\pm0.31}$	78.24$_{\pm1.99}$
	M3	59.47$_{\pm0.28}$	65.04$_{\pm0.08}$	65.11$_{\pm5.72}$	80.01$_{\pm0.11}$	78.71$_{\pm1.61}$	**85.36**$_{\pm2.95}$	85.18$_{\pm0.73}$	88.25$_{\pm2.49}$	96.53$_{\pm1.40}$	**89.40**$_{\pm1.48}$	**74.23**$_{\pm5.76}$	**97.52**$_{\pm0.58}$	80.40$_{\pm1.93}$
	M4	48.64$_{\pm1.04}$	39.45$_{\pm0.04}$	72.93$_{\pm3.25}$	80.00$_{\pm0.01}$	61.44$_{\pm0.63}$	39.26$_{\pm2.12}$	68.49$_{\pm0.03}$	64.74$_{\pm1.82}$	65.20$_{\pm1.53}$	68.70$_{\pm0.69}$	51.34$_{\pm2.86}$	88.06$_{\pm0.03}$	62.35$_{\pm1.17}$
	M5	48.58$_{\pm0.20}$	32.71$_{\pm0.04}$	71.02$_{\pm1.27}$	80.00$_{\pm0.01}$	81.54$_{\pm3.07}$	85.71$_{\pm1.51}$	81.04$_{\pm2.07}$	85.48$_{\pm3.93}$	96.65$_{\pm1.17}$	80.52$_{\pm1.78}$	66.21$_{\pm4.48}$	96.85$_{\pm0.45}$	71.41$_{\pm1.87}$
	M6	52.54$_{\pm0.41}$	62.70$_{\pm3.00}$	65.66$_{\pm2.34}$	80.61$_{\pm0.16}$	71.23$_{\pm3.62}$	71.17$_{\pm8.00}$	84.43$_{\pm2.50}$	83.05$_{\pm1.66}$	95.84$_{\pm2.91}$	84.45$_{\pm4.46}$	64.33$_{\pm3.05}$	94.86$_{\pm1.14}$	75.91$_{\pm2.77}$
	M7	58.05$_{\pm0.81}$	62.61$_{\pm0.06}$	56.78$_{\pm2.30}$	80.86$_{\pm0.12}$	78.02$_{\pm4.03}$	75.01$_{\pm1.20}$	77.97$_{\pm1.27}$	86.94$_{\pm4.40}$	94.39$_{\pm2.43}$	81.18$_{\pm1.79}$	57.86$_{\pm5.98}$	96.58$_{\pm0.34}$	75.52$_{\pm2.06}$
	M8	59.35$_{\pm0.46}$	68.99$_{\pm0.06}$	**88.20**$_{\pm0.01}$	83.35$_{\pm0.10}$	81.47$_{\pm2.82}$	80.14$_{\pm1.38}$	87.04$_{\pm1.26}$	88.30$_{\pm1.42}$	97.22$_{\pm0.04}$	85.49$_{\pm1.18}$	66.09$_{\pm3.68}$	97.40$_{\pm0.74}$	81.92$_{\pm1.10}$
RF	M1	**58.91**$_{\pm0.76}$	**81.89**$_{\pm0.20}$	**84.66**$_{\pm0.54}$	**87.63**$_{\pm0.79}$	**81.97**$_{\pm0.89}$	82.16$_{\pm2.15}$	84.38$_{\pm1.56}$	88.20$_{\pm0.98}$	91.81$_{\pm1.58}$	**93.87**$_{\pm4.63}$	69.26$_{\pm2.02}$	**97.79**$_{\pm0.66}$	**83.54**$_{\pm1.40}$
	M2	56.03$_{\pm0.61}$	78.70$_{\pm0.11}$	81.68$_{\pm1.14}$	86.07$_{\pm0.50}$	81.56$_{\pm1.73}$	80.26$_{\pm1.78}$	83.98$_{\pm1.05}$	89.98$_{\pm0.05}$	93.48$_{\pm0.71}$	91.84$_{\pm1.29}$	**74.42**$_{\pm4.15}$	97.29$_{\pm0.32}$	82.94$_{\pm1.12}$
	M3	55.78$_{\pm0.80}$	75.59$_{\pm0.12}$	78.94$_{\pm0.75}$	84.13$_{\pm0.48}$	81.24$_{\pm1.00}$	75.47$_{\pm1.93}$	85.90$_{\pm1.74}$	90.40$_{\pm1.98}$	95.06$_{\pm2.81}$	91.13$_{\pm0.80}$	69.59$_{\pm3.43}$	97.01$_{\pm1.37}$	81.69$_{\pm1.43}$
	M4	50.65$_{\pm0.19}$	51.31$_{\pm0.67}$	79.80$_{\pm1.25}$	72.91$_{\pm0.50}$	63.31$_{\pm1.86}$	51.60$_{\pm3.60}$	56.01$_{\pm0.90}$	70.21$_{\pm2.10}$	61.34$_{\pm0.76}$	68.96$_{\pm3.99}$	49.19$_{\pm4.41}$	84.46$_{\pm2.26}$	63.31$_{\pm1.87}$
	M5	50.25$_{\pm0.32}$	52.17$_{\pm0.42}$	82.92$_{\pm0.58}$	82.27$_{\pm0.34}$	80.81$_{\pm2.81}$	78.78$_{\pm5.59}$	85.26$_{\pm1.75}$	88.75$_{\pm3.50}$	93.73$_{\pm3.10}$	93.88$_{\pm2.66}$	62.82$_{\pm3.48}$	97.35$_{\pm0.73}$	76.52$_{\pm1.89}$
	M6	57.47$_{\pm0.62}$	68.09$_{\pm3.21}$	83.44$_{\pm1.36}$	83.75$_{\pm0.27}$	78.96$_{\pm5.06}$	**86.37**$_{\pm3.09}$	**86.05**$_{\pm2.71}$	**92.61**$_{\pm1.82}$	**96.08**$_{\pm1.57}$	93.46$_{\pm1.52}$	69.00$_{\pm4.90}$	96.41$_{\pm0.94}$	82.64$_{\pm2.26}$
	M7	56.57$_{\pm0.70}$	79.05$_{\pm0.25}$	82.38$_{\pm0.61}$	79.73$_{\pm0.59}$	80.42$_{\pm2.84}$	77.86$_{\pm4.11}$	76.61$_{\pm3.05}$	88.67$_{\pm1.28}$	94.06$_{\pm3.47}$	94.90$_{\pm1.83}$	62.07$_{\pm5.19}$	97.18$_{\pm0.72}$	80.79$_{\pm2.05}$
	M8	56.62$_{\pm1.03}$	77.20$_{\pm0.03}$	83.94$_{\pm0.84}$	86.91$_{\pm0.62}$	79.53$_{\pm3.89}$	79.52$_{\pm1.89}$	83.59$_{\pm1.17}$	91.06$_{\pm1.59}$	93.23$_{\pm2.54}$	93.70$_{\pm2.47}$	65.35$_{\pm4.84}$	97.29$_{\pm0.92}$	82.33$_{\pm1.82}$

To sum up, the 3WDO model, compared with the LOSSA for sparse streaming feature completion, reduces the decision risk. Thus, the 3WDO model is more suitable with missing data rates from 0.1 to 0.9.

3) Accuracy Analysis with $\zeta = 0.1$: We take the missing rate of 10% as an example of detailed analysis. Fig. S2 in the supplementary file shows the mean number of selected features in the supplementary file. The predictive accuracy of three classifiers against these competing algorithms is recorded in Table XIII.

The Friedman Test: The *ρ-value* of the Friedman test on KNN, SVM, RF, and the mean number of selected features are 1.6959e-10, 7.1241e-11, 4.2587e-12, and 5.9840e-05, respectively. Thus, there are significant differences between the 3WDO model and different algorithms, with 10% missing data.

The Mean Number of Selected Features: For the 3WDO model, the mean number of selected features is stable for different sparse data sets. It is stable between 5 and 10. However, some algorithms, such as the SAOLA, significantly differ in the number of selected features on different data sets. And we have the following findings.

1) Some algorithms select the maximum feature while getting lower accuracy than the 3WDO, for example, Fast-OSFS on D2. The redundancy analysis is insufficient due to many redundant features being retained.
2) And some algorithms select the least; for example, the SFS-FI only selects one feature in some data sets. The reason may be that the algorithm cannot select all the essential features of 10% missing data sets and only select the first features on some data sets, causing the loss of critical information.
3) The 3WDO makes adequate relevance and redundancy analysis; meanwhile, different thresholds exist for additional features, and it does not lose important information, ensuring select fewer features to get the highest accuracy.

The Classification Accuracy of Selected Features: From Table XIII, we find that the 3WDO model performs better than its rivals on eight of the twelve data sets in the case of SVM and KNN and seven data sets with RF. In addition, the 3WDO model gets the lowest average ranks and higher average accuracy of three classifiers than other algorithms. Then, we have the following observations.

1) *The 3WDO Versus the Fast-OSFS:* The 3WDO is higher than Fast-OSFS on most competing data sets. The Fast-OSFS handle the incomplete streaming features relying on zero-filling and employs certain relevance and redundancy analysis, causing misclassification and inferior accuracy performance.
2) *The 3WDO Versus the SAOLA:* The SAOLA only considers the feature relationships between two features. By contrast, with the full use of the LFA model and 3WD, the 3WDO can always select the critical features on the fly while having better predictive accuracy.
3) *The 3WDO Versus the SFS-FI:* The sparse streaming features cause some feature interaction information loss. The SFS-FI performs the worst on accuracy. Therefore, it cannot handle sparse streaming features well.

1) *The 3WDO Versus the OSSFS-DD:* The OSSFS-DD only has an average accuracy of around 0.43 on D2 and 0.97 on D12. The unsteady average accuracy on the sparse data sets is that the OSSFS-DD also adopts the 3WD in relevance analysis. Still, the threshold is only dynamically updated by calculation. Unlike the automatic updating of the threshold in this article, the updating threshold value method of the 3WDO is universal. This demonstrates that the 3WDO is more stable than the OSSFS-DD in accuracy.
2) *The 3WDO Versus the RHOFS:* RHOFS uses fuzzy sets to select features, which improves the accuracy, but the effect of feature selection is not too good when the data set is missing. Therefore, the 3WDO algorithm needs to complete the data with the LFA model before selecting features.
3) *The 3WDO Versus the ANOHFS:* ANOHFS employs adaptive neighborhood rough set to select features, which enhances the efficiency, but the performance of feature selection becomes relatively poor when dealing with sparse streaming features. Therefore, it is effective for the 3WDO algorithm to use the LFA model to complete missing values.
4) *The 3WDO Versus the LOSSA:* Considering complete sparse streaming features, the accuracy of the LOSSA is second only to the 3WDO. The results show that the LFA model completes sparse streaming features is effective. Though the sparse matrix is completed, the LOSSA uses certain relevance and redundancy analysis, which causes lower average accuracy than the 3WDO. Conversely, the 3WDO modifies the 3WD to reduce the decision risk caused by uncertainty and get the highest accuracy.

To sum up, compared with OSFS and OS^{2}FS models, sparse streaming features completed by the LFA model generally avoids information loss and improve performance. Hence, the 3WDO and LOSSA perform best on sparse streaming features. Meanwhile, the features selected by the 3WDO have higher accuracy than the LOSSA, demonstrating that three-way relevance analysis with minimum decision risk contributes to better performance.

C. Error Analysis of Missing Value Completion (R.Q. 3)

Analyze the error of completing missing values in the initial stage, i.e., the accuracy of preprocessing sparse streaming feature matrix, and confirm its convergence capability. The root mean squared error (RMSE) is used as an evaluation index, as depicted in (20). In (20), the data set completion error training curve is illustrated for missing rates of 0.1, 0.3, 0.5, 0.7, and 0.9, respectively. Fig. S3 in the supplementary file shows that the RMSE of data completion gradually converges as the number of iterations increases on D1-12 in the supplementary file, and a completion matrix with a lower error than the original value can be quickly obtained. However, as the missing rate increases, the accuracy of predicting missing values decreases. The reason is that with a higher missing rate, the known data of training is reduced, resulting in insufficient information to construct the complete spare streaming features. As shown in the accuracy of feature selection, even when the missing rate is high, the accuracy of the feature model for the 3WDO model remains higher than others on most datasets

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{f'_{n,j} \in \Lambda_U} \left(f'_{n,j} - p_{n,k} q_{j,k}\right)^2}{\Lambda_U}} \tag{22}$$

TABLE XIV
AVERAGE ACCURACY OF SELECTED FEATURES WITH NONRANDOMLY DISTRIBUTED MISSING DATA

Models/Datasets	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	Average	^Rank
M1	**57.59**$_{\pm 0.64}$	82.52$_{\pm 0.14}$	**85.31**$_{\pm 0.67}$	**86.37**$_{\pm 0.45}$	**83.90**$_{\pm 3.33}$	**82.18**$_{\pm 2.72}$	**86.87**$_{\pm 1.16}$	**91.87**$_{\pm 1.38}$	**96.90**$_{\pm 1.91}$	**93.29**$_{\pm 1.52}$	**67.83**$_{\pm 4.71}$	**98.12**$_{\pm 0.28}$	**84.40**$_{\pm 1.58}$	**1.17**
M2	51.56$_{\pm 0.75}$	83.29$_{\pm 0.20}$	73.51$_{\pm 5.97}$	85.24$_{\pm 0.56}$	77.54$_{\pm 2.83}$	64.40$_{\pm 3.02}$	83.98$_{\pm 1.12}$	84.38$_{\pm 1.89}$	95.70$_{\pm 1.87}$	89.61$_{\pm 1.86}$	58.08$_{\pm 3.31}$	97.73$_{\pm 0.63}$	78.75$_{\pm 2.00}$	4.00
M3	51.57$_{\pm 0.63}$	70.51$_{\pm 0.21}$	72.71$_{\pm 5.72}$	80.60$_{\pm 0.28}$	73.89$_{\pm 1.18}$	75.55$_{\pm 2.97}$	85.16$_{\pm 1.47}$	83.76$_{\pm 2.12}$	95.87$_{\pm 1.52}$	89.82$_{\pm 2.14}$	59.35$_{\pm 3.68}$	96.98$_{\pm 0.64}$	77.98$_{\pm 1.88}$	4.42
M4	49.14$_{\pm 0.50}$	67.38$_{\pm 0.22}$	71.67$_{\pm 5.14}$	76.26$_{\pm 0.43}$	79.26$_{\pm 3.63}$	36.92$_{\pm 3.68}$	62.00$_{\pm 1.63}$	80.42$_{\pm 2.20}$	54.02$_{\pm 1.72}$	68.44$_{\pm 3.15}$	37.66$_{\pm 3.05}$	86.19$_{\pm 0.99}$	64.11$_{\pm 2.20}$	7.33
M5	50.93$_{\pm 0.72}$	**86.10**$_{\pm 0.15}$	76.71$_{\pm 0.97}$	83.11$_{\pm 0.39}$	78.43$_{\pm 3.75}$	81.17$_{\pm 4.09}$	81.89$_{\pm 1.57}$	77.07$_{\pm 3.42}$	84.65$_{\pm 2.87}$	79.77$_{\pm 1.68}$	57.38$_{\pm 3.84}$	92.81$_{\pm 0.67}$	77.50$_{\pm 2.01}$	5.00
M6	49.54$_{\pm 0.83}$	72.88$_{\pm 0.22}$	73.91$_{\pm 1.23}$	84.98$_{\pm 0.37}$	70.15$_{\pm 4.00}$	51.08$_{\pm 2.06}$	78.41$_{\pm 1.75}$	83.75$_{\pm 2.71}$	96.80$_{\pm 1.77}$	87.38$_{\pm 2.12}$	61.40$_{\pm 4.49}$	95.69$_{\pm 0.75}$	75.50$_{\pm 1.86}$	5.17
M7	50.37$_{\pm 0.77}$	64.30$_{\pm 0.15}$	75.84$_{\pm 1.33}$	78.82$_{\pm 0.19}$	70.28$_{\pm 1.55}$	58.21$_{\pm 2.79}$	69.27$_{\pm 1.50}$	78.77$_{\pm 2.17}$	86.63$_{\pm 3.69}$	82.93$_{\pm 1.89}$	64.30$_{\pm 0.15}$	91.84$_{\pm 0.64}$	72.63$_{\pm 1.40}$	6.17
M8	56.67$_{\pm 0.67}$	80.52$_{\pm 0.1}$	85.27$_{\pm 0.67}$	85.43$_{\pm 0.31}$	80.00$_{\pm 3.92}$	80.43$_{\pm 2.33}$	82.53$_{\pm 0.87}$	91.67$_{\pm 1.78}$	96.48$_{\pm 1.52}$	92.83$_{\pm 1.04}$	65.55$_{\pm 4.66}$	95.53$_{\pm 0.75}$	82.74$_{\pm 1.56}$	2.75

^ The Average rank.

TABLE XV
AVERAGE ACCURACY OF SELECTED FEATURES VARYING WITH DIFFERENT INITIAL TEMPERATURE

initT/Datasets	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	Average	^Rank
10	**58.84**$_{\pm 0.59}$	77.38$_{\pm 0.21}$	86.35$_{\pm 0.37}$	**87.01**$_{\pm 0.48}$	83.98$_{\pm 1.16}$	**83.74**$_{\pm 2.23}$	**86.55**$_{\pm 1.64}$	90.82$_{\pm 1.76}$	**96.44**$_{\pm 1.36}$	92.60$_{\pm 3.50}$	**66.68**$_{\pm 2.94}$	97.97$_{\pm 0.43}$	84.03$_{\pm 1.39}$	1.83
20	57.89$_{\pm 0.78}$	**79.05**$_{\pm 0.22}$	86.45$_{\pm 0.62}$	85.81$_{\pm 0.32}$	81.28$_{\pm 3.88}$	75.07$_{\pm 4.40}$	85.02$_{\pm 1.63}$	**91.00**$_{\pm 1.38}$	94.29$_{\pm 2.55}$	**93.60**$_{\pm 1.32}$	65.32$_{\pm 3.46}$	97.57$_{\pm 0.64}$	82.70$_{\pm 1.77}$	3.25
30	58.07$_{\pm 0.79}$	77.57$_{\pm 0.75}$	**86.51**$_{\pm 0.63}$	86.59$_{\pm 0.61}$	83.87$_{\pm 2.42}$	79.77$_{\pm 4.59}$	84.24$_{\pm 1.14}$	90.12$_{\pm 1.62}$	95.11$_{\pm 1.87}$	92.86$_{\pm 1.85}$	62.21$_{\pm 3.87}$	**98.08**$_{\pm 0.60}$	82.92$_{\pm 1.73}$	2.75
40	57.03$_{\pm 0.62}$	73.67$_{\pm 0.41}$	86.28$_{\pm 0.53}$	86.92$_{\pm 0.40}$	84.17$_{\pm 4.21}$	76.32$_{\pm 5.03}$	85.31$_{\pm 1.93}$	89.67$_{\pm 3.00}$	92.29$_{\pm 2.51}$	91.25$_{\pm 0.88}$	66.01$_{\pm 3.13}$	97.60$_{\pm 0.54}$	82.21$_{\pm 1.93}$	3.67
50	58.05$_{\pm 0.68}$	72.32$_{\pm 1.50}$	86.04$_{\pm 1.21}$	84.95$_{\pm 0.33}$	**87.25**$_{\pm 3.21}$	79.47$_{\pm 7.12}$	85.96$_{\pm 0.97}$	89.69$_{\pm 1.99}$	95.98$_{\pm 1.94}$	91.47$_{\pm 1.51}$	65.98$_{\pm 3.71}$	97.31$_{\pm 0.45}$	82.87$_{\pm 2.05}$	3.50

D. Accuracy Analysis With Nonrandomly Distributed Missing Data (R.Q. 4)

This section studies the accuracy comparison between 3WDO and other algorithms on missing data with nonrandom distribution. Due to external factors such as equipment aging, it is assumed that the probability of the entire feature column being missing gradually increases from 10% to 90% over time. Table XIV records the accuracy of feature selection on the missing data with nonrandom distribution. It can be seen that on most of the data, 3WDO and LOSSA perform better, mainly because the LFA model also has a relatively good effect in complementing the missing data with nonrandom distribution. Moreover, 3WDO is better than LOSSA, which proves that 3WDO further reduces the negative impact of uncertain feature selection.

E. Parameter Analysis (R.Q. 5)

The p*arameter analysis* aims to find the optimal parameters for the 3WDO model. Through a systematic evaluation process, various parameter values are tested, and their impact on the model's performance is assessed. By conducting multiple trials with different parameter configurations and analyzing the resulting outcomes, the most suitable parameter values are identified. The experimental findings provided valuable insights into the optimal parameter settings that can enhance the model's overall effectiveness.

1) Analysis on Parameter initT: This experiment analyzes the effect of different values of initial temperature *initT*. We fix the missing data rate ζ at 0.1, $\lambda = 0.01$, $h = 10$, and $L = 5$, and apply the 3WDO model on the different values of *initT* $= 10$, 20, 30, 40, and 50 on these twelve data sets, respectively. Table XV records the average accuracy with three classifiers of different values. The ρ*-value* of the Friedman test on average accuracy with three classifiers is 1.4052e-07. Thus, there is a significant difference between different values of *initT*. The features selected by the 3WDO of other values have extra accuracy. These results demonstrate that the *initT* affects the accuracy of feature selection, so finding the appropriate value is crucial. And then, we should choose the most suitable value for the 3WDO model. The *initT* $= 10$ performs better than other values on twelve data sets, getting lower average ranks and higher average accuracy. Therefore, *initT* $= 10$ is more suitable for this experiment. The ρ-value of the Wilcoxon signed-ranks test of *initT* $= 10$ compared with other values on average accuracy are 0.0461, 0.0225, 0.0012, and 0.0100. Therefore, the effect of *initT* $= 10$ is better. In summary, it is not that the higher the value of *initT*, the better the effect of feature selection; It is best to choose the *initT* value according to the accuracy. Besides, with the consideration of time complexity, we set *initT* $= 10$ to evaluate the proposed 3WDO model.

2) Analysis on Parameter λ: This experiment aims to analyze the impact of the parameter λ on the accuracy of 3WDO. The study investigates the effects of five λ values (0.01, 0.02, 0.03, 0.04, and 0.05) on experimental results. The parameters are set as $\zeta = 0.1$, $h = 10$, *initT* $= 10$, and $L = 5$. Table S(II) in the supplementary file presents the experimental outcomes of 3WDO using five different λ values (0.01, 0.02, 0.03, 0.04, and 0.05) in the supplementary file. Furthermore, the Friedman test indicates a significant influence of the λ value on feature selection results, with a ρ**-value** of 1.4052e-07. The results in Table S(II) in the supplementary file show that higher λ values yield worse performance. Specifically, $\lambda = 0.01$ achieves the highest average value, indicating superior performance. This is attributed to regularization improving missing value completion and reducing overfitting. However, excessively large λ values can decrease the accuracy of feature selection. The ρ**-value** of the Wilcoxon signed-ranks test of $\lambda = 0.01$ and 0.02, 0.03, 0.04, and 0.05 are 0.0461, 0.0225, 0.0012, and 0.0100, respectively, indicating that $\lambda = 0.01$ is better than other different values. Therefore, selecting the most suitable λ value for the 3WDO model is crucial, with the recommended value being 0.01.

3) Analysis on Parameter h: When new features are received, they should be temporarily stored in the buffer.

Once the data in the buffer reaches its capacity, the missing data should be completed first before proceeding with the feature selection process. Does the size of the buffer impact the accuracy of feature selection? The parameters are set as $\zeta = 0.1$, $initT = 10$, $\lambda = 0.01$, and $L = 5$. Table S(III) in the supplementary file displays the average accuracy of feature selection for 12 datasets with varying h values of 10, 20, 30, 40, and 50 in the supplementary file. The results in Table S(III) in the supplementary file show significant fluctuations in average classification accuracy with changes in parameters particularly for datasets D6 and D10. The Friedman test examines the accuracy of all data, and the ρ**-value** is 1.2191e-07. The experimental results show that the performance of feature selection is sensitive to parameter h. The larger the value, the lower the accuracy of feature selection. Therefore, the value of parameter h can be adjusted to suit 3WDO to obtain better classification accuracy. They are setting $h = 10$ to yield the highest average accuracy and optimal outcomes. The Wilcoxon signed-ranks test for $h = 10$ compared to 20, 30, 40, and 50 yield ρ**-value** of 0.0015, 4.8828e-04, 0.0046, and 0.0012, respectively, with $h = 10$ showing the best feature selection performance.

4) Analysis on Parameter L: To complete the sparse data, the original data needs to be

mapped into two low-dimensional matrices, which raises the question of whether the dimensions of the matrices affect the results of feature selection. In this experiment, all other parameters remain unchanged, and the dimensions are set to 5, 10, 15, and 20 to investigate the impact of dimensions on the experimental results, as presented in Table S(IV) in the supplementary file. The difference in feature selection accuracy across various dimensions is evident. In addition, the ρ**-value** of the Friedman test in Table S(IV) in the supplementary file is 9.3080e-06, indicating that dimension indeed impacts feature selection accuracy. Therefore, selecting the most suitable dimension value for the 3WDO model is crucial. In order to achieve the highest accuracy, this article opts for a dimension of 5. The ρ**-value** of the Wilcoxon signed-ranks test of 5 with 10, 15, and 20 are 2.4414e-04, 2.4414e-04, and 0.0081, respectively. This demonstrates that setting the dimension to 5 in the 3WDO model results in improved feature selection effectiveness.

VI. Conclusion

In this article, we study the problem of the OS^{2}FS model and propose a three (3)-way decision-incorporated OS^{2}FS (3WDO) model. Its main idea is to adopt the LFA model to complete the sparse matrix on the fly, reducing errors caused by massive missing data. Besides, with the help of 3WD, carrying out uncertain relevance and redundancy analysis with minimum decision risk. The 3WDO algorithm is tested on twelve classified data sets. The experimental results verify that the 3WDO algorithm can effectively improve the ability of the OS^{2}FS algorithm. Note that some parameters of 3WDO, including the size of the buffer matrix and regularization coefficient, need time-consuming and tedious manual tuning. Furthermore, the parameters can be made adaptive by evolutionary algorithms or other feasible frameworks that can improve the practicability of 3WDO.

It is noted that simulated annealing is used to find the threshold, and whether other adaptive methods are more suitable for it needs further study. Meanwhile, the LFA model combined with other methods, such as the Laplacian, will get higher accuracy. In our further work, we will study the technique of completing missing data and combine various ways to select online sparse streaming features.

References

[1] F. Yao, Y. Ding, S. Hong, and S.-H. Yang, "A survey on evolved LoRa-based communication technologies for emerging Internet of Things applications," *Int. J. Netw. Dyn. Intell.*, vol. 1, no. 1, pp. 4–19, 2022.

[2] S. Ramírez-Gallego et al., "An information theory-based feature selection framework for big data under apache spark," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 48, no. 9, pp. 1441–1453, Sep. 2018.

[3] K. Yuan, D. Miao, W. Pedrycz, H. Zhang, and L. Hu, "Multi-granularity data analysis with zentropy uncertainty measure for efficient and robust feature selection," *IEEE Trans. Cybern.*, vol. 55, no. 2, pp. 740–752, Feb. 2025, doi: 10.1109/TCYB.2024.3499952.

[4] K. Yuan, D. Miao, W. Pedrycz, W. Ding, and H. Zhang, "Ze-HFS: Zentropy-based uncertainty measure for heterogeneous feature selection and knowledge discovery," *IEEE Trans. Knowl. Data Eng.*, vol. 36, no. 11, pp. 7326–7339, Nov. 2024.

[5] K. Yuan, D. Miao, Y. Yao, H. Zhang, and X. Zhao "Feature selection using zentropy-based uncertainty measure," *IEEE Trans. Fuzzy Syst.*, vol. 32, no. 4, pp. 2246–2260, Apr. 2024.

[6] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Comput. Electr. Eng.*, vol. 40, no. 1, pp. 16–28, 2014.

[7] S. Alelyani, J. Tang, and H. Liu, "Feature selection for clustering: A review," in *Data Clustering*. Boca Raton, FL, USA: Chapman Hall/CRC, 2018, pp. 29–60.

[8] P. P. Kundu and S. Mitra, "Feature selection through message passing," *IEEE Trans. Cybern.*, vol. 47, no. 12, pp. 4356–4366, Dec. 2017.

[9] Y. Yang, D. Chen, H. Wang, and X. Wang, "Incremental perspective for feature selection based on fuzzy rough sets," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 3, pp. 1257–1273, Jun. 2018.

[10] X. Xue, M. Yao, and Z. Wu, "A novel ensemble-based wrapper method for feature selection using extreme learning machine and genetic algorithm," *Knowl. Inf. Syst.*, vol. 57, no. 2, pp. 389–412, 2018.

[11] H. Liu and H. Motoda, *Computational Methods of Feature Selection*. London, U.K.: Chapman Hall, 2007.

[12] B. Xue, M. Zhang, W. N. Browne, and X. Yao, "A survey on evolutionary computation approaches to feature selection," *IEEE Trans. Evol. Comput.*, vol. 20, no. 4, pp. 606–626, Aug. 2016.

[13] J. Ni, H. Fei, W. Fan, and X. Zhang, "Automated medical diagnosis by ranking clusters across the symptom-disease network," in *Proc. IEEE Int. Conf. Data Min.*, 2017, pp. 1009–1014.

[14] G. Ditzler, J. LaBarck, J. Ritchie, G. Rosen, and R. Polikar, "Extensions to online feature selection using bagging and boosting," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 9, pp. 4504–4509, Sep. 2018.

[15] X. Wu, K. Yu, W. Ding, H. Wang, and X. Zhu, "Online feature selection with streaming features," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 5, pp. 1178–1192, May 2013.

[16] K. Yu, X. Wu, W. Ding, and J. Pei, "Scalable and accurate online feature selection for big data," *ACM Trans. Knowl. Discov. Data*, vol. 11, no. 2, pp. 1–39, 2016.

[17] M. B. Badsha et al. "Imputation of single-cell gene expression with an autoencoder neural network," *Quant. Biol.*, vol. 8, no. 1, pp. 78–94, 2020.

[18] A. Idri, H. Benhar, J. L. Fernández-Alemán, and I. Kadi, "A systematic map of medical data preprocessing in knowledge discovery," *Comput. Methods Programs Biomed.*, vol. 162, pp. 69–85, Aug. 2018.

[19] D. Wu, Y. He, X. Luo, and M. Zhou, "A latent factor analysis-based approach to online sparse streaming features selection," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 52, no. 11, pp. 6744–6758, Nov. 2022.

[20] F. Chen, D. Wu, J. Yang, and Y. He, "Online sparse streaming feature selection with uncertainty," in *Proc. IEEE Int. Conf. Netw., Sens. Control (ICNSC)*, Shanghai, China, 2022, pp. 1–6.

[21] Y. Yao, "The geometry of three-way decision," *Appl. Intell.*, vol. 51, pp. 6298–6325, Jan. 2021.

[22] S. Perkins and J. Theiler, "Online feature selection using grafting," in *Proc. 20th Int. Conf. Mach. Learn.*, 2003, pp. 592–599.

[23] J. Zhou, D. P. Foster, R. A. Stine, and L. H. Ungar, "Streamwise feature selection," *J. Mach. Learn. Res.*, vol. 7, no. 67, pp. 1861–1885, 2006.

[24] P. Zhou, N. Wang, and S. Zhao, "Online group streaming feature selection considering feature interaction," *Knowl.-Based Syst.*, vol. 226, Aug. 2021, Art. no. 107157.

[25] P. Zhou, P. Li, S. Zhao, and X. Wu, "Feature interaction for streaming feature selection," *IEEE Trans. Neural. Netw. Learn. Syst.*, vol. 32, no. 10, pp. 4691–4702, Oct. 2021.

[26] P. Zhou, S. Zhao, Y. Yan, and X. Wu, "Online scalable streaming feature selection via dynamic decision," *ACM Trans. Knowl. Discov. Data*, vol. 16, no. 5, pp. 1–20, 2022.

[27] P. Zhou, X. Hu, P. Li, and X. Wu, "Online streaming feature selection using adapted neighborhood rough set," *Inf. Sci.*, vol. 481, pp. 258–279, May 2019.

[28] P. Zhou, X. Hu, P. Li, and X. Wu, "OFS-density: A novel online streaming feature selection method," *Pattern Recognit.*, vol. 86, pp. 48–61, Feb. 2019.

[29] C. Luo, S. Wang, T. Li, H. Chen, J. Lv, and Z. Yi, "RHDOFS: A distributed online algorithm towards scalable streaming feature selection," *IEEE Trans. Parallel Distrib. Syst.*, vol. 34, no. 6, pp. 1830–1847, Jun. 2023.

[30] T. Shu, Y. Lin, and L. Guo, "Online hierarchical streaming feature selection based on adaptive neighborhood rough set," *Appl. Soft Comput.*, vol. 152, Feb. 2024, Art. no. 111276.

[31] Y. Koren, R. Bell, and C. Volinsky, "Matrix-factorization techniques for recommender systems," *IEEE Comput.*, vol. 42, no. 8, pp. 30–37, Aug. 2009.

[32] Z. Zhang, W. Jiang, F. Li, M. Zhao, B. Li, and L. Zhang, "Structured latent label consistent dictionary learning for salient machine faults representation-based robust classification," *IEEE Trans. Ind. Informat.*, vol. 13, no. 2, pp. 644–656, Apr. 2017.

[33] J.-D. Zhang, C.-Y. Chow, and J. Xu, "Enabling kernel-based attribute-aware matrix factorization for rating prediction," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 4, pp. 798–812, Apr. 2017.

[34] X. Luo, Z. Liu, S. Li, M. Shang, and Z. Wang, "A fast non-negative latent factor model based on generalized momentum method," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 51, no. 1, pp. 610–620, Jan. 2021.

[35] M. Gong, X. Jiang, H. Li, and K. C. Tan, "Multiobjective sparse non-negative matrix factorization," *IEEE Trans. Cybern.*, vol. 49, no. 8, pp. 2941–2954, Aug. 2019.

[36] X. Luo, Z. Wang, and M. Shang, "An instance-frequency-weighted regularization scheme for non-negative latent factor analysis on high-dimensional and sparse data," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 51, no. 6, pp. 3522–3532, Jun. 2021.

[37] X. Luo, Z. Wang, and M. Shang, "An instance-frequency-weighted regularization scheme for non-negative latent factor analysis on high dimensional and sparse data," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 51, no. 6, pp. 3522–3532, Jun. 2021.

[38] D. Wu, X. Luo, M. Shang, Y. He, G. Wang, and M. Zhou, "A deep latent factor model for high-dimensional and sparse matrices in recommender systems," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 51, no. 7, pp. 4285–4296, Jul. 2021.

[39] Y. Yao, "Three-way decision: An interpretation of rules in rough set theory," in *Proc. Int. Conf. Rough Sets Knowl. Technol.*, 2009, pp. 642–649.

[40] D. Liu, "The effectiveness of three-way classification with interpretable perspective," *Inf. Sci.*, vol. 567, pp. 237–255, Aug. 2021.

[41] Y. Yao, "Three-way decision and cognitive computing," *Cogn. Comput.*, vol. 8, no. 4, pp. 543–554, 2016.

[42] Y. Yao, "Three-way decision and granular computing," *Int. J. Approx. Reason.*, vol. 103, pp. 107–123, Dec. 2018.

[43] Y. Zhang and J. Yao, "Gini objective functions for three-way classifications," *Int. J. Approx. Reason.*, vol. 81, pp. 103–114, Feb. 2017.

[44] Q. Zhang, D. Xia, and G. Wang, "Three-way decision model with two types of classification errors," *Inf. Sci.*, vol. 420, pp. 431–453, Dec. 2017.

[45] X. D. Yue, X. F. Chen, D. Q. Miao, and H. Fujita, "Fuzzy neighborhood covering for three-way classification," *Inf. Sci.*, vol. 507, pp. 795–808, Jan. 2020.

[46] D. Guo, W. Xu, Y. Qian, and W. Ding, "Fuzzy-granular concept-cognitive learning via three-way decision: Performance evaluation on dynamic knowledge discovery," *IEEE Trans. Fuzzy Syst.*, vol. 32, no. 2, pp. 1409–1423, Mar. 2024.

[47] D. Guo, C. Jiang, R. Sheng, and S. Liu, "A novel outcome evaluation model of three-way decision: A change viewpoint," *Inf. Sci.*, vol. 607, pp. 1089–1110, Aug. 2022.

[48] Y. Yao, "An outline of a theory of three-way decision," in *Proc. Int. Conf. Rough Sets Current Trends Comput.*, 2012, pp. 1–17.

[49] W. Xu, D. Guo, Y. Qian, and W. Ding, "Two-way concept-cognitive learning method: A fuzzy-based progressive learning," *IEEE Trans. Fuzzy Syst.*, vol. 31, no. 6, pp. 1885–1899, Jun. 2023.

[50] W. Xu, D. Guo, J. Mi, Y. Qian, K. Zheng, and W. Ding, "Two-way concept-cognitive learning via concept movement viewpoint," *IEEE Trans. Neural. Netw. Learn. Syst.*, vol. 34, no. 10, pp. 6798–6812, Oct. 2023.

[51] J. M. Pea, "Learning Gaussian graphical models of gene networks with false discovery rate control," in *Proc. Eur. Conf. Evol. Comput., Mach. Learn. Data Min. Bioinf. (EvoBIO)*, 2008, pp. 165–176.

[52] Y. Yao, "The superiority of three-way decision in probabilistic rough set models," *Inf. Sci.*, vol. 181, no. 6, pp. 1080–1096, 2011.

[53] Y. Yao, "Three-way decision with probabilistic rough sets," *Inf. Sci.*, vol. 180, no. 3, pp. 341–353, 2010.

[54] T. Tlili and S. Krichen, "A simulated annealing-based recommender system for solving the tourist trip design problem," *Expert Syst. Appl.*, vol. 186, Dec. 2021, Art. no. 115723.

[55] X. Hu, P. Zhou, P. Li, J. Wang, and X. Wu, "A survey on online feature selection with streaming features," *Front. Comput. Sci.*, vol. 12, no. 3, pp. 479–493, 2018.

Ruiyang Xu (Graduate Student Member, IEEE) received the B.S. degree in information and computing science from Qufu Normal University, Qufu, China, in 2020, and the Electronic Information degree in computer science from the School of Computer Science and Information Engineering, Harbin Normal University, Harbin, China, in 2022. She is currently pursuing the joint Ph.D. degree in computer science with the school of computer science and technology, Chongqing University of Posts and Telecommunications, Chongqing, united training by the Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences, Chongqing, China.

Her research interests include streaming feature selection, big data analysis, and algorithm design for large-scale data applications.

Di Wu (Member, IEEE) received the Ph.D. degree in computer application technology from the Chongqing Institute of Green and Intelligent Technology (CIGIT), Chinese Academy of Sciences (CAS), China, in 2019.

He then joined CIGIT, CAS, China. He is currently a Professor with the College of Computer and Information Science, Southwest University, Chongqing, China. He has over 70 publications, including 17 IEEE TRANSACTIONS papers and several conference papers on AAAI, ICDM, WWW, and IJCAI. His research interests include machine learning and data mining.

Prof. Wu is serving as an Associate Editor for the *Neurocomputing* and *Frontiers in Neurorobotics*. His homepage: https://wudi1986.github.io/Homepage/.

Renfang Wang received the B.S. degree in computer application from the Huazhong University of Science and Technology, Wuhan, China, in 2000, the M.S. degree in computer application technology from the Henan University of Science and Technology, Luoyang, China, in 2004, and the Ph.D. degree in computer science and technology from Zhejiang University, Hangzhou, China, in 2008.

He joined the College of Big Data and Software Engineering, Zhejiang Wanli University, Ningbo, China, in 2008, as a Professor of Computer Science and Technology. His current research interests include big data analysis and computer vision.

Xin Luo (Fellow, IEEE) received the B.S. degree in computer science from the University of Electronic Science and Technology of China, Chengdu, China, in 2005, and the Ph.D. degree in computer science from Beihang University, Beijing, China, in 2011.

He is currently a Professor of Data Science and Computational Intelligence with the College of Computer and Information Science, Southwest University, Chongqing, China. He has authored or co-authored over 300 papers, including over 160 IEEE Transactions/Journal papers in the areas of his interests. His research interests include artificial intelligence, data science, and neural networks.

Prof. Luo was the recipient of the Outstanding Associate Editor Award from IEEE Access in 2018, the IEEE/CAA Journal of Automatica Sinica in 2020, and from the IEEE Transactions on Neural Networks and Learning Systems in 2022. He is currently serving as an Associate Editor for IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, and *IEEE/CAA Journal of Automatica Sinica*. His Google Scholar homepage is at the link of https://scholar.google.com/citations?user=hyGlDs4AAAAJ&hl=zh-TW.